The Children Activity Scanning Tool (CAST): a methodological study of an instrument to measure physical activity engagement levels of children in school playgrounds and similar environments: a study of the instrument’s development through four health promotion projects in NSW Australia

Avigdor Zask
Southern Cross University

Publication details
Zask, A 2009, 'The Children Activity Scanning Tool (CAST): a methodological study of an instrument to measure physical activity engagement levels of children in school playgrounds and similar environments: a study of the instrument's development through four health promotion projects in NSW Australia', PhD thesis, Southern Cross University, Lismore, NSW.
Copyright A Zask 2009
The Children Activity Scanning Tool (CAST):
A methodological study of an instrument to measure physical activity engagement levels of children in school playgrounds and similar environments

A study of the instrument’s development through
Four health promotion projects in NSW Australia

A thesis presented to the School of Health & Human Sciences Southern Cross University
Lismore, NSW, Australia

Submitted in fulfilment of the requirement of the degree of Doctor of Philosophy

Avigdor Zask
Grad. Dip. Soc. Ec. (University of Western Sydney)
I certify that the work presented in this thesis is, to the best of my knowledge and belief, original, except as acknowledged in the text, and that the material has not been submitted, either in whole or in part, for a degree at this or any other university.

I acknowledge that I have read and understood the University's rules, requirements, procedures and policy relating to my higher degree research award and to my thesis. I certify that I have complied with the rules, requirements, procedures and policy of the University (as they may be from time to time).

Name: Avigdor Zask

Signature:…………………………………………………………………..

Date: 22 August 2009

Acknowledgements

I would like to thank my supervisors Associate Professor Steve Kermode and Dr Eric van Beurden for their input, support and patience, and to Dr Lyndon Brooks from the SCU Graduate Research College for advice and guidance regarding methodology and statistical analyses.

I would like to acknowledge the North Coast Area Health Service Health Promotion Team for allowing me to use data obtained while working for the team and granting me study leave to research and write the thesis.

To all the Health promotion colleagues and research assistants in all projects without whom I would not have had the data.

Thanks to all the participating schools, their staff and students for accommodating the research.

Last, but not least, thanks to my partner Suzanne Rienits for her support over the years and her proof reading of the manuscript.
Abstract

Thesis aim

This thesis aims to analyse and critically evaluate the Children Activity Scanning Tool (CAST), which measures children’s physical activity (PA) engagement levels in school playgrounds and similar environments, focusing on CAST reliability and validity using data from four health promotion projects in the Northern Rivers (Move it Groove it (MIGI) project), Illawarra, Central Coast and Dubbo areas in NSW, Australia.

The overall research question is whether physical activity engagement levels of large numbers of children can be validly and reliably measured in school playgrounds and similar environments using CAST.

Methodology

CAST was used to scan school playgrounds and similar environments, in which unstructured rest, play and other physical activities take place. Scanning was undertaken by a team of trained observers who concurrently scanned a pre-defined area of the playground and recorded the number of children engaged in assigned physical activity intensity categories.

Criterion validity was measured in three projects comparing the total number of children counted in separate physical activity categories to an independent count of the number of children in the playground using Pearson correlation coefficients. Data obtained during instrument development were used to determine criterion validity in two projects when compared to videotapes using Pearson correlation, Cronbach alphas and 95% CI around mean engagement levels.

Data obtained during development and training, as well as project field data were used for reliability testing. Cronbach alpha, Pearson correlation, factor analysis and 95% CI around mean engagement levels were used to measure inter-rater reliability.
Data from multi-level regression models obtained in three projects were used to calculate intra class correlation (ICC) and composite reliabilities of scan and break level physical activity engagement.

**Results**

Criterion validity measures comparing sums of categories counts to independent counts of all children yielded Pearson correlation coefficients of 0.97, 0.81 and 0.96 in the Illawarra, Central Coast and Dubbo projects respectively. Criterion validity measures comparing field and video scores yielded Pearson correlation coefficients of 0.47, 0.52 and 0.56 respectively in the Move it Groove it (MIGI) project, and in the Illawarra study Cronbach alphas of 0.98 and 0.85, and a Pearson correlation of 0.93. There were no significant differences between the mean numbers of children who were engaged in a physical activity level when 95% CI were used except for one category in the MIGI study.

Inter-rater reliability estimates in all projects were high or very high. There were no significant differences between the mean numbers of children engaged in a physical activity level when 95% CI were used. Almost 86% (12/14) of the standardised Cronbach alpha estimates were equal to or greater than 0.93. Principal component analysis values ranged from 76%-98%. ICCs obtained from field data yielded composite reliabilities of 0.77-0.94.

**Discussion**

**Conclusion**
It was concluded that CAST was a valid and reliable instrument to measure physical activity engagement levels of children in school playgrounds.

**Recommendations**
The two or three categories versions CAST should be used as they produced the highest quality data when used to observe large groups in well-defined playgrounds. Further research is needed to validate other variables (eg equipment), validate a one observer CAST version, and test the feasibility of using children as observers.
Table of contents

Abstract ..................................................................................................................... iii
Thesis aim .............................................................................................................. iii
Methodology ........................................................................................................... iii
Results ................................................................................................................... iv
Discussion ............................................................................................................. iv
Conclusion..................................................................................................... iv
Recommendations......................................................................................... iv

Table of contents....................................................................................................... v

1. Introduction ............................................................................................................ 1
   Research questions............................................................................................... 2
   Conceptual framework for the thesis .................................................................... 3
   Thesis structure....................................................................................................... 4
Theoretical context ..................................................................................................... 5
   The epidemiology and health effects of childhood and adult obesity ....................... 5
   Benefits of physical activity for children’s health ...................................................... 6
   Tracking of overweight and obesity from childhood to adulthood ......................... 6
   History of childhood obesity as a developing public health issue ................................. 7
   Mechanisms involved in the development of obesity ................................................. 8
   The effects of food intake, physical activity and sedentariness ..................................... 10
   Children’s physical activity and sedentary behaviours .............................................. 10
   Factors influencing physical activity in children ....................................................... 13
   Guidelines for adequate physical activity (PA) ........................................................ 13
   Rationale for promoting children’s physical activity in schools .................................... 15
   Health promotion and children’s physical activity .................................................... 17
Methodological context ........................................................................................... 21
   Measuring children’s PA engagement during school breaks ....................................... 21
   The history and development of CAST .................................................................. 23
   Projects in which CAST has been used ................................................................. 27
   Measurement reliability and validity in behavioural and social science .................... 29
Summary of the introduction chapter ....................................................................... 33

2. Methods............................................................................................................... 35
   Field instrument administration procedures ......................................................... 36
   Sample and settings in which CAST was developed and used ................................... 41
   Move It Groove It (MIGI) project – project data ..................................................... 41
   Illawarra project .................................................................................................... 41
   Central Coast project ............................................................................................. 41
   Dubbo project ....................................................................................................... 41
   Testers and training ............................................................................................... 42
3. Results ................................................................................................................ 56

Sample and settings overview .................................................................................. 57

Project field data - sample and settings .................................................................. 57

MIGI project ................................................................................................................. 57

Illawarra project .......................................................................................................... 57

Central Coast project .................................................................................................. 58

Dubbo project .............................................................................................................. 58

CAST reliability - samples and results ...................................................................... 58

MIGI project ................................................................................................................. 58

Consensus estimation regarding category identification ............................................. 58

Consistency estimation regarding number of children in each activity level ............... 58

Inter-rater reliability as part of the video validation study .......................................... 59

Consistency estimation using ICCs – project field data ............................................. 59

Illawarra project .......................................................................................................... 60

Consistency estimation re number of children in each activity level ........................... 60

Consistency estimation using ICCs – project field data ............................................. 60
List of Appendices ........................................................................................................ 112
Appendix 1: CAST training manual (from Central Coast project) ............................................. 113
Appendix 2: Preventive Medicine paper regarding the MIGI project playground component ......... 123
Appendix 3: Health Promotion Journal of Australia brief report regarding the Illawarra project ....... 131
Appendix 4: Physical activity engagement results of the Illawarra project ....................................... 135
Appendix 5: Physical activity engagement results of the Central Coast project ................................. 139
Appendix 6: MIGI project evaluation summary report ................................................................. 141
Appendix 7: British Journal of School Nursing paper about the use of CAST in another project ....... 145
List of tables
Table 1: A comparison between 3 different methods of measuring physical activity among children in school playtime ..........................................................................................................................................22
Table 2: Summary of methods to estimate reliability used in each study ..............................................................49
Table 3: Means and 95% confidence intervals of the six CAST raters’ scores in the MIGI project for four activity categories ............................................................................................................59
Table 4: Means and 95% confidence intervals of the five CAST raters’ scores in the Illawarra project for three activity categories ......................................................................................................60
Table 5: Means and 95% confidence intervals of the seven CAST raters’ scores in the Central Coast project for two activity categories and the overall number of children ........................................................................61
Table 6: Means and 95% confidence intervals of the three CAST raters’ scores in the Dubbo project for two activity categories and the overall number of children ........................................................................63
Table 7: Summary of results – reliability estimates used in each study ...................................................................64
Table 8: Mean number of children engaged in physical activity categories and 95% confidence intervals by activity levels in the four studies ..............................................................................65
Table 9: Means and 95% confidence intervals of field and video scores in the MIGI project validation study ..........................67
Table 10: Means and 95% confidence intervals of field and video scores in the Illawarra project validation study ........................................................................................................68
Table 11: Parameter estimates and standard errors from variance components and significant effects only for MVPA and VPA in the MIGI project ..................................................................................70
Table 12: Parameter estimates and standard errors from variance components and significant effects only for MVPA and VPA in the Illawarra project ..............................................................................72
Table 13: Parameter estimates and standard errors from variance component and significant effects only models for MVPA ............................................................................................................74
Table 14: Summary of different versions of CAST, their features, limitations and types of reliability and validity measures used .............................................................................................................89

List of Figures
Figure 1: Conceptual framework of the thesis ..................................................................................................................4
Figure 2: A sample map of a school playground area with two scanning areas .............................................................37
Figure 3: A sample map of a school playground area with one scanning area ..................................................................38
1. Introduction
This thesis aims to analyse and critically evaluate the development of a direct observation instrument, which measures children’s physical activity engagement levels in school playgrounds. The instrument is the Children Activity Scanning Tool (CAST), and the thesis documents its evolution from its design and usage by the North Coast Health Promotion Team in the Move it Groove it (MIGI) project\textsuperscript{1, 2}, through its use in three other projects.

The focus of this thesis is on methodological aspects of CAST, specifically in relation to instrument reliability and validity and also to advantages and disadvantages of the different versions of the instrument.

**Research questions**

The overall research question is whether physical activity engagement levels of large numbers of children can be measured in school playgrounds and similar environments using a direct observation instrument. A more specific research hypothesis is that CAST can validly and reliably measure children’s physical activity levels in school playgrounds. This can be broken down into three operational null hypotheses, as follows:

1. Physical activity levels of large numbers of children in school playgrounds cannot be measured.
2. CAST is not a valid instrument to measure children’s physical activity levels in school playgrounds.
3. CAST is not a reliable instrument to measure children’s physical activity levels in school playgrounds.

Part of this methodological analysis and critique involved comparison between the different versions of CAST, which used different numbers of observers to measure a different number of physical activity categories. This involved initial assessment of reliability and validity of its usage with multiple observers and physical activity categories. Subsequently, data from two other smaller studies were used to assess CAST reliability, when performed by only two observers and measuring only two physical activity categories (active and sedentary).
The development and testing of CAST is inseparable from the issue of childhood obesity since increasing children’s physical activity is a very important strategy to arrest the rise in childhood obesity and generally improve children’s health and development. The thesis therefore focuses on factors, which are amenable to change through health promotion interventions. In this sense it is ‘applied research’ and is primarily based on fieldwork and research performed by the author in his work for the North Coast Area Health Service, and in collaboration with health promotion workers from other Area Health Services. Consequently, the introduction section outlines in some detail the evidence regarding societal and environmental factors influencing childhood obesity and, where relevant, refers to evidence concerned with genetic and other biological factors.

The methods and results sections are concerned mainly with measuring CAST reliability and validity, and not whether the interventions, in which CAST has been used, resulted in changes in children’s physical activity levels. Nevertheless the results and findings of the actual project evaluations that have utilized CAST, and have also been the field test sites for its development, are included as appendices (Appendices 2, 4 & 5).

**Conceptual framework for the thesis**

There are three major threads to this thesis – contexts, methods and outcomes. Within these threads are embedded the various analytic and critical perspectives, which are brought to bear on the topic. This framework was used to address the research question systematically and to allow each of the operational hypotheses to be tested against the available data. The thesis framework is represented in figure 1.
Thesis structure
While the conceptual framework outlined above was used to guide the thesis, the Introduction-Methods-Results-Discussion (IMRD) structure was still used. See Table of Contents for details.
Theoretical context

This section will review the epidemiology and health effects of obesity in general and childhood obesity in particular. It will explore the mechanisms involved in obesity, in terms of the environmental factors of food intake, physical activity, sedentary behaviours and the relationship between them. It will then overview the development of guidelines for children’s recommended daily physical activity levels and the rationale for promoting physical activity in schools. Physical activity promotion will then be placed in the broader context of the field of Health Promotion.

The focus will then shift to physical activity related research and in particular to research measuring children’s physical activity engagement during school breaks. This will lead to an outline of the development of CAST, followed by a description of the projects in which CAST has been used. The section will conclude with an overview of the concepts of reliability and validity and details regarding common reliability and validity types.

The epidemiology and health effects of childhood and adult obesity

Obesity in general and childhood obesity in particular, has become a major public health concern in the last 2 decades.3-17 It is conservatively estimated that in the US, poor diet and physical inactivity contributes to 15% of the burden of preventable disease9, 10 and that obesity contributes to 112,000 -280,000 deaths annually.5

Ebbeling et al. state that:
“Obesity in childhood causes high blood pressure, increased blood clotting tendency, dyslipidaemia, chronic inflammation, endothelial dysfunction, and hyperinsulinaemia. This clustering of cardiovascular disease risk factors, known as the insulin resistance syndrome, has been identified in children as young as 5 years of age”.8

Focusing on the metabolic syndrome, Weiss et al. report that: “The prevalence of the metabolic syndrome is high among obese children and adolescents, and it increases with worsening obesity”.16 More than 60% of overweight children have at least one other risk factor for cardiovascular disease, such as raised blood pressure,
hyperlipidaemia, and hyperinsulinaemia, and more than 20% have 2 or more risk factors.\textsuperscript{7}

The prevalence of type 2 diabetes among adolescents is rising and now accounts for over 30% of all new diagnoses in parts of the US.\textsuperscript{7, 19} This condition was rarely seen among adolescents in the past; hence its old name was ‘Adult Onset Diabetes’.\textsuperscript{20}

There is also evidence that lifestyle modification programs of three months duration significantly improve risk factors for metabolic syndrome and insulin resistance in overweight youth. This adds to the evidence regarding the relationship between childhood obesity and risk factors, but also points to the potential public health benefit of preventive programs.\textsuperscript{21}

There are also psychosocial consequences of childhood obesity. Obese children and adolescents often suffer early and systematic social discrimination.\textsuperscript{22}

**Benefits of physical activity for children’s health**

As well as maintaining a healthy body weight, the World Health Organisation states in a fact-sheet that general health benefits of regular adequate physical activity for children are: developing healthy musculoskeletal tissues (i.e. bones, muscles and joints), developing a healthy cardiovascular system (i.e. heart and lungs), and developing neuromuscular awareness (i.e. coordination and movement control).\textsuperscript{23}

The fact-sheet further states that:

“Physical activity has also been associated with psychological benefits in young people by improving their control over anxiety and depression. Similarly, participation in physical activity can assist in the social development of young people by providing opportunities for self-expression, building self-confidence, social interaction and integration”.\textsuperscript{23}

**Tracking of overweight and obesity from childhood to adulthood**

Further evidence indicates that childhood obesity is related to increased adult mortality and that overweight tendencies, as well as cardiovascular disease and metabolic syndrome markers, track moderately well from adolescence to adulthood.\textsuperscript{24} There are significant associations between being overweight in
childhood and both overall mortality and ischemic heart disease mortality after 57 years. The risk of being obese as an adult is greater for children over 3 years of age who are overweight regardless of parental weight.

As well as the effects on physical health, it is also well documented that childhood obesity adversely effects psycho social health. There is also evidence regarding the effect of children’s involvement in physical activity and sport on psycho-social outcomes including its effectiveness as a crime prevention strategy.

**History of childhood obesity as a developing public health issue**

The issue of childhood obesity first appeared in medical literature in the late 1950s. One of the first people to write about the seriousness of childhood obesity and the importance of treating obese children in the clinic was Hoffman in 1957. In 1974, an editorial in the Lancet alerted readers to the emerging phenomenon and its consequences. While childhood overweight and obesity was a phenomenon worth noticing in the 1970’s, it seems that at least in the UK, its prevalence stayed stable until the mid 1980’s and then started rising sharply. Chinn & Rona reported that: “Little change was found in the prevalence of overweight or obesity from 1974 to 1984. From 1984 to 1994 overweight increased from 5.4% to 9.0% in English boys (3.6%, 95% CI 2.3-5.0%) and from 6.4% to 10.0% in Scottish boys (3.6%, 1.9-5.4%). Values for girls were 9.3% to 13.5% (4.1%, 2.4-5.9%) and 10.4% to 15.8% (5.4%, 3.2-7.6%), respectively. The prevalence of obesity increased correspondingly, reaching 1.7% (English boys), 2.1% (Scottish boys), 2.6% (English girls), and 3.2% (Scottish girls”).

Covering a shorter period, yet showing a similar trend, overweight and obesity levels among NSW children and youth have increased significantly from 1985 to 1997 and continued increasing at a slower rate from 1997 to 2004. There was a 1.9 fold increase in the rate of overweight youth (9.4%-18.4%), a 5.5 fold increase in the rate of obese youth (1.4%-7.7%), and a 2.4 fold increase in the rate of combined overweight and obese youth (10.8%-26.1%) between 1985 and 2004.

The proportion of children who are overweight or obese has increased in the last 20 years in both developed and developing countries. In developing countries the rates of childhood overweight and obesity have grown between 1.5 and 3 fold over
the last 10-20 years and are higher among children of lower socio economic status (SES). The rates of childhood overweight and obesity, in developing countries in the same period, have grown 3 to 4 fold and are higher among children of higher SES. This growth in childhood obesity in developing countries is attributed to adoption of western lifestyles and urbanisation. In most of these countries both childhood overweight and obesity and childhood malnutrition are public health concerns.

In view of the health consequences for the individual, these increases are likely to have major public health consequences. Some researchers warn that, if these trends continue unmitigated, life expectancy in developed countries may stop rising, and that obese people will enjoy fewer disability-free life years and experience higher rates of diabetes, hypertension, heart disease and osteoporosis.

While the measurement of physical activity among children, which is described in this thesis, was initiated as part of health promotion units’ response to childhood obesity prevention, it is important to recognise that increasing physical activity has health benefits among people who are overweight or obese even if their weight status does not change. There is very strong evidence for the beneficial health effects of regular adequate physical activity on physiological processes and indicators associated with preventing the most common chronic diseases. These include muscle and liver insulin sensitivity, muscle glucose uptake and utilization, and overall glycemic control, resulting in improved lipid profile, lower blood pressure, and a positive effect on thromboembolic state. Regular adequate physical activity also improves balance and has positive effects on musculoskeletal conditions.

**Mechanisms involved in the development of obesity**

Obesity is essentially the consequence of continuous greater energy intake than expenditure. Accumulating fat is a biological mechanism, which has evolved to store energy in times of adequate food supply, to support survival in times of famine.

“Obesity is clearly the result of the interaction between genetic and environmental factors and is a complex multifactorial condition developing from interactive influences of numerous factors - social, behavioural, physiological, metabolic, cellular, and molecular.”
Studies of individuals with a wide range of Body Mass Index (BMI – Weight in kilograms/height in meters squared)\textsuperscript{39}, with data obtained on their parents, siblings, and spouses suggest that about 25 to 40 percent of the individual differences in body mass, body fat or waist girth may depend on genetic factors.\textsuperscript{38, 40} Predisposition to common forms of obesity is probably influenced by numerous susceptibility genes, accounting for variations in energy requirements, fuel utilization, muscle metabolic characteristics, and taste preferences.\textsuperscript{40} In their review paper “Obesity – is it a genetic disorder?” Loos and Bouchard answer the question positively for only 5% of obesity cases and a larger percentage of the severely obese, so only a very small proportion of obese people are obese due only to genetic factors. They state that: “the obesity epidemic we are facing today occurred only over the past 3 decades and can clearly not be explained by changes in our genome”. \textsuperscript{40}

The main thrust of Loos and Bouchard’s answer to the above question is a model of genetic susceptibility to obesogenic environments. They conclude that major behavioural, environmental and societal changes would be needed to arrest the spread of the obesity epidemic in developed countries and around the world.\textsuperscript{40}

Even at the physiological level the mechanisms appear to be more complex than previously thought. A 2008 study by Kechagias et al.\textsuperscript{41} investigated high calorie fast food intake following the documentary film “Supersize Me”.\textsuperscript{42} A new Scientist report cites the study to highlight the ability of some people to burn off excess energy they have taken in through producing excess heat in their bodies.\textsuperscript{43} However, even those who have gained less weight gained five percent or more of their body weight (with the exception of one subject out of eighteen). Most study participants also had pathological levels of the liver enzyme ALT following the relatively short period in which the study was undertaken (four weeks) while controls’ levels did not change.\textsuperscript{41}

There has been some debate and disagreement regarding one of these environmental influences, dietary intake. Stubbs and Lee state that “Recent data from Australia, the United States and Europe show increased self-reported energy intake associated with obesity, in contrast to earlier suggestions that the obesity epidemic has occurred despite minimal or no increase in per capita energy intake from food.” \textsuperscript{44} They conclude, however, that “Both physical activity and nutrition must
be addressed to reduce the prevalence of obesity and improve the health of Australians. The evidence is therefore clear that inadequate physical activity is one of the important factors explaining the rising prevalence of obesity.

The effects of food intake, physical activity and sedentariness
As identified in the last section, the three behavioural factors affecting overweight and obesity levels through disruption of the energy balance are increased food intake, decreased physical activity, and increase in sedentary pursuits.

It is clear that children today have more options to consume more calories, contained in easily available energy dense foods, than in the past. An American national study found that 30% of children in the total sample, reported consuming fast food on a typical day, and that eating fast food had an adverse effect on dietary quality in ways that plausibly could increase risk for obesity. Other studies concur with these findings and emphasise the effects of high fat and especially high fructose intake on energy metabolism and the central regulation of appetite leading to insulin resistance, overweight and obesity among children.

There is evidence regarding the important role sweetened drinks play in childhood overweight and obesity, by adding to the total energy intake, as well as reducing children’s Calcium intake through reduced milk intake. There is also evidence that reducing sweet drinks intake among British youth through a health promotion intervention had the desired effect on children’s health. The intervention resulted in arresting of weight gain in intervention children (0.2% increase) while control children continued to put on weight (7.5% increase). In a pilot study of childhood obesity prevention in preschools project in Northern NSW, Australia, a significant association was found between the number of sweet drinks a child consumed and the probability of being obese.

Children’s physical activity and sedentary behaviours
It is clear that increasing physical activity would be crucial to arrest the rise in adult obesity prevalence. However, it is not clear whether children’s energy expenditure is significantly lower than it was in the past as there are no available data on children’s physical activity patterns before the 1980’s to compare with current data. While it seems plausible that western children in the 1950’s and 1960’s spent more
time in incidental physical activity, including commuting to school, as well as more
time in unsupervised play, and less time in sedentary pursuits, the evidence in
support of this hypothesis is complex.

There are a number of studies on children’s levels of physical activity. British studies
confirmed that the modern sedentary lifestyle starts in early childhood. One study
found that the time young children were engaged in moderate and vigorous physical
activity (MVPA) constituted less than 5% of their waking hours.\(^{52}\) A large study
comparing data from 34 countries found that the percentage of youth aged 10-16
years, who were active for 60 or more minutes on 5 or more days per week, ranged
between 19.3% in France and 49.5% in the US. There is other evidence that
substantial numbers of children are not adequately active\(^{11, 12, 53, 54}\).

In New South Wales (NSW) Australia, there was no decline in children’s physical
activity levels between 1985 and 2004. The 2004 NSW Schools Physical Activity
and Nutrition Survey had a component that could be compared with the 1985
Australian Health and Fitness Survey.\(^{27}\) Students were asked to recall the
frequency and duration of time spent in activities such as walking and cycling to
school, physical education classes, and sport. Surprisingly, there were significant
increases in the proportion of boys and girls who were engaged in MVPA for 60
minutes or more per day. For boys, these increases were 15% and 20% for year 8
and year 10 students respectively, while among girls these increases were around
25% and 20%.\(^{27}\) The same study also found significant decreases in the number of
children walking or cycling to school, but active commuting was included in overall
reported physical activity, so it seems that youth in NSW may have increased their
engagement in MVPA despite or possibly as a response to reducing their cycling
and walking to school.\(^{27}\)

An important limitation of this study was the use of children and adolescents self-
reported physical activity as the data source. Societal attitudes to and children’s
awareness of childhood obesity and physical activity may have changed between
1985, 1997 and 2004, so the reliability of these estimates is unknown and they
should be interpreted with caution.
While these findings are very encouraging for NSW in terms of youth physical activity levels, the same study also confirmed that, like in other countries, overweight and obesity levels among Australian children and youth have increased significantly from 1985 to 1997 and continued increasing at a slower rate from 1997 to 2004. There was a 1.9 fold increase in the rate of overweight youth (9.4%-18.4%), a 5.5 fold increase in the rate of obese youth (1.4%-7.7%), and a 2.4 fold increase in the rate of combined overweight and obese youth (10.8%-26.1%) between 1985 and 2004. 27

While Booth et al.'s findings 27 point to the need to address children's food intake, time spent in sedentary pursuits might also play a part in the childhood obesity equation. 7, 8, 27 Time spent in such pursuits, especially watching TV, has been found to be correlated with increased risk of being overweight and obese among children. 55 Robinson previously found that when an intervention reduced children's TV watching time, children in the intervention group had statistically significant relative decreases in body mass index. 56 The author of this thesis has found significant association between the number of evening meals eaten while the TV was on and preschool children’s waist girth. 57

There is also evidence regarding association between increased energy dense food consumption and decreased fruit and vegetable consumption among children in families with high TV viewing habits. 57, 58 Coon et al. found that this association remained significant after adjustment for socio economic status (SES). 58 The amount of time spent watching television by children is an indirect measure of sedentary behaviour. 59 Some programs have attempted to reduce sedentary behaviours such as watching television and found that active play simultaneously increased. 59 There are indications that television viewing promotes the development and maintenance of overweight by making vigorous activities less likely and by increasing caloric intake while watching television. 60 Compounding this problem, is the fact that many advertisements during prime time TV for children appear to be for energy dense foods. 8, 61

Another intriguing area of evidence has linked sedentary behaviour to decreased daily energy expenditure. Levine et al. argue that humans expend energy through
purposeful exercise and through changes in posture and movement that are associated with the routines of daily life (fidgeting and being restless) called non-exercise activity thermogenesis (NEAT). They found that “Obese individuals appear to exhibit an innate tendency to be seated for 2.5 hours per day more than sedentary lean counterparts. If obese individuals were to adopt the lean ‘NEAT-o-type’, they could potentially expend an additional 350 kcal per day.” Levine is involved in a project that reconfigures the classroom environment so that lessons are held in spaces such as halls or gyms where children have no seats, and access their computer stations either whilst standing or sitting on the floor.

Factors influencing physical activity in children
A 2000 review of factors influencing physical activity in primary school children and youth found that for primary school students the following correlates were significant: “sex (male), parental overweight status, physical activity preferences, intention to be active, perceived barriers (inverse), previous physical activity, healthy diet, program/facility access, and time spent outdoors”. Regarding youth, it was found that:

“Variables that were consistently associated with adolescents' physical activity were sex (male), ethnicity (white), age (inverse), perceived activity competence, intentions, depression (inverse), previous physical activity, community sports, sensation seeking, sedentary after school and on weekends (inverse), parent support, support from others, sibling physical activity, direct help from parents, and opportunities to exercise”.

Guidelines for adequate physical activity (PA)
The complexity of children’s physical activity levels and emerging evidence about the increase in childhood overweight and obesity have been reflected in the recommendations for adequate children and youth physical activity in the last decade or so.

The 1994 Physical activity guidelines for adolescents consensus statement recommended at least 30 minutes of accumulated daily engagement in Moderate or vigorous physical activity (MVPA) with three additional sessions weekly of vigorous physical activity of at least 20 minutes duration. For elementary school age children, the 1998 American guidelines were that children should accumulate at
least 30 to 60 minutes of physical activity on most or all days of the week, and should be encouraged to accumulate up to several hours per day.\(^6\)

In 2003 these guidelines were updated.\(^6\) The authors recommended that children should accumulate at least 60 minutes and up to several hours of physical activity each day, that children should participate in several bouts of physical activity lasting 15 minutes or more, and that extended periods of inactivity (more than two hours) should be discouraged especially during daytime. The Australian recommendations for children and youth 5-18 years of age are similar:

“Children and young people should participate in at least 60 minutes (and up to several hours) of moderate- to vigorous-intensity physical activity every day. Children and young people should not spend more than 2 hours a day using electronic media for entertainment (eg computer games, Internet, TV), particularly during daylight hours”.\(^6\)

The authors of the 1998 American guidelines conceded in their statement that more research is needed in this field and that:

“When possible, these guidelines are based on research; however, many are based on behavioural theory and standards for exemplary practice in physical education, exercise science, health education, and public health. More research is needed on the relationship between physical activity and health among young people, the relationship between physical activity during childhood and adolescence and that during adulthood, the determinants of physical activity among children and adolescents, and the effectiveness of school and community programs promoting physical activity among young people”.\(^6\)

While it makes sense to consider clear and achievable cut-off levels for health promotion interventions, some researchers argue that the physical activity levels recommended in these guidelines may be insufficient to produce the desired health effects among children. In 2006, a large international study on clustered risk factors for cardiovascular disease (CVD) and metabolic syndrome among children and youth concluded that:
“Physical activity levels should be higher than the current international guidelines of at least 1 hour per day of physical activity of at least moderate intensity, to prevent clustering of cardiovascular disease risk factors.” ⁶⁹

The authors of the above study found that the bottom 3 quintiles of their sample had raised risk factors for CVD in all analyses. The mean time spent in MVPA by the children in the fourth quintile was 116 min per day in 9-year-old and 88 min per day in 15-year-old children. These findings led the authors to suggest a revision of international guidelines for children’s recommended daily PA levels to 90 minutes per day of PA of at least moderate intensity to prevent clustering of cardiovascular disease risk factors. ⁶⁹ An earlier study by Epstein et al., which reviewed 26 earlier studies that used heart rate monitoring with youth, speculated that youth likely need 120-150 minutes of daily physical activity engagement for health benefits. ⁷⁰

In another development, cut-off points of recommended numbers of daily steps (as measured by pedometers) were related to healthy body composition for boys and girls aged 6-12 years. These were calculated following a study of American, Australian, and Swedish children. The selected cut points for steps per day were 12,000 steps/day for girls and 15,000 steps/day for boys. ⁷¹ However, a major limitation of using pedometers, which the authors cite, is that pedometers are not capable of recording physical activity intensity or bouts of activity. ⁷¹

**Rationale for promoting children’s physical activity in schools**

Following the evidence outlined so far, it is clear that promoting physical activity and discouraging excessive sedentary behaviours among children is a public health priority.

On most days children spend up to half their waking hours either in school or getting to and from school. Schools are therefore important as settings, which could provide opportunities for children to be active. Recognising the role of schools in achieving required PA levels for children, the NSW (Australia) Board of Studies recommended that children in NSW schools should have the opportunity to engage in 120 minutes of PA during the school week. This should include activity in Physical Education (PE) classes as well as activity during recess and participation in school sports activities. ⁷²
Children who do not get adequate opportunity to be active in school do not compensate by being more active outside the school. This may be surprising, but may be explained by the amount of time spent by children watching television, playing video games and using computers. In a national US survey, 67% of children watched television more than 2 hours a day and 26% watched TV for more than 4 hours a day. In a survey of parents of preschool children in the North Coast of NSW, 22% reported that their child spent more than 2 hours in screen based activities on the day preceding the survey. In rural areas in Australia, children often spend long periods of time commuting to and from school on a daily basis, reducing the time available for them to be physically active. Thus, schools may be an important setting for children to be active.

Recess and break periods comprise about a sixth of the school day and provide the greatest opportunity for children to be active during school hours. In terms of preparing children for life long engagement in PA, the school playground provides an environment more similar to unstructured adult recreational environments than PE classes.

Opportunities to engage in MVPA during Physical Education (PE) classes vary greatly and depend on two factors: the frequency of PE classes and the opportunities to be active during PE classes. Levels of both these factors have been shown to be less than desirable. A number of interventions to increase the time children spend being physically active during PE lessons produced significant results. However, these increases, while encouraging, only added a small number of minutes to children’s total daily physical activity. PE lessons have other important functions in terms of child development and health, especially the acquisition of movement skills in preparation for participation in physical activity, sport and exercise. Better movement skills are significantly, though weakly, associated with habitual physical activity.

There is little evidence regarding how much of children’s total daily physical activity is accumulated during school recess. Ridgers et al. reviewed the literature regarding children’s physical activity levels during playtime and stated that playtime was found
to contribute between 14 and 35.5 minutes of engagement in MVPA, or up to a half of the recommended hour of daily MVPA. Wickel et al. studied the contribution of youth sport, recess and PE lessons to daily physical activity using accelerometers. They found that, during weekdays without youth sport activities, recess accounted for 26% and 19% of the daily minutes of activity for boys and girls respectively, while PE lessons accounted for 8% and 9%. On days with youth sport activities, recess accounted for 18% and 23% of daily minutes of activity for boys and girls respectively, and PE lessons accounted for less than 5%. Ridgers and Stratton reported that boys and girls spent 26 and 20 minutes respectively engaged in MVPA during recess. They add that: “These results suggest that recess can make a worthwhile contribution to the recommended 60 minutes of MVPA per day”. However, their study used heart rate monitors, which may reflect other factors such as stress and excitement as well as MVPA engagement. Ridgers et al. conclude in another paper that it would be a realistic target for schools to achieve the engagement of children in MVPA for 40% of recess times in schools.

Health promotion and children’s physical activity
To provide the reader with an understanding of the rationale for, and application of, the instruments presented in this thesis, it is important to explain the context within which it was developed. This context is the field of health promotion. Aspects of the field of health promotion relevant to physical activity promotion will be reviewed in this section.

Health Promotion is a relatively new discipline, which emerged in the early 1970’s, although it is also related to the public health drive from the turn of the last century. Health Promotion uses a dynamic multi-disciplinary approach to achieve preventive population health goals. The World Health Organisation (WHO) auspiced Alma Ata declaration acknowledged that health is more than the absence of disease, and expanded the individual perspective of health to encompass the idea that people are affected by their social, economic and natural environments.

In 1981, following the Declaration of Alma Ata, the WHO developed the Global Strategy for Health for All by the Year 2000. The major themes of the strategy were: equity in health, health promotion, enhancing preventative activities through
primary health care settings, cooperation between government, community and the private sector, and the need to increase community participation.

In the 1980’s WHO acknowledged that the Alma Ata Declaration provided a theoretical base as well as an ethical imperative to develop a primary health care approach, but there was no identifiable framework for action. The first International Conference on Health Promotion was held in Ottawa, Canada and resulted in the Ottawa Charter for health promotion, which provided an action framework for health promotion implementation.97

Health was again defined as a resource for life rather than the absence of disease and its physical, mental, emotional and spiritual aspects were acknowledged. Socio-political and environmental prerequisites for health such as peace, shelter, education, food, income, a stable ecosystem, sustainable resources, social justice and equity were listed as a requirement for improvement in health. Health promotion was defined as “the process of enabling people to increase control over and to improve their health”. The charter also articulated five domains of action for health promotion practitioners: Create supportive environments (for health), build healthy public policies, strengthen community action, reorient health services (to focus more ‘upstream’, i.e. on prevention), and develop personal skills.97

These action domains or strategies reflected the eclectic influences on the health promotion field, like social ecology, community development, organisational development, behavioural psychology, social marketing, education, environmental and social justice activism, as well as biomedical sciences.92, 94, 98-100 These strategies have been incorporated into health promotion planning processes, so contemporary health promotion projects often include a combination of most or all of these strategies.101

Further refinement of the Ottawa charter strategies in health promotion practice have included the approaches of inter-sectoral collaboration and coalition building, and the intention to improve health promotion program sustainability through community, organisational and individual capacity building, as well as policy development and implementation.99, 101, 102 Another strong influence on health promotion practice has been the evidence based practice movement. Health
promotion practitioners are increasingly informed by evidence about health issues as well as evidence regarding the efficacy of interventions to address them.\textsuperscript{103, 104}

All of the general health promotion principles outlined above can potentially be harnessed in the promotion of children’s physical activity. The evidence regarding the magnitude of the childhood obesity issue, and the realisation that it may be a new ‘epidemic’, has provided the impetus for health promotion researchers to gather more specific data on the factors contributing to the problem.

Evidence such as that presented in the preceding sections of this thesis has been used to plan health promotion programs as health and education departments have realised the importance of addressing the issue of childhood obesity in general and children’s physical activity in particular. Consequently, during the 1990’s and into the 2000’s there have been many physical activity promotion programs in schools, mostly multi-strategic, and mostly conducted in partnership with education sectors and other related organisations.\textsuperscript{1, 2, 84, 105-110}

Despite the growing interest and investment in children’s physical activity promotion, some public health advocates in the US argue that in comparison to other areas of health promotion, the science of physical activity promotion at a population level is still in its infancy. Furthermore, Yancey et al. claim that physical activity promotion lacks public health infrastructure.\textsuperscript{111}

Considering such claims, it seems that NSW is ahead of the field in providing at least a strategic direction for physical activity promotion. Advocacy from leading health promotion and public health researchers and administrators regarding government action on childhood obesity prevention has resulted in the 2002 NSW Childhood Obesity Summit. The Summit initiated an integrated whole-of-government approach and a government action plan to address childhood obesity.\textsuperscript{112-115} This action plan addressed physical activity promotion through a number of strategies including environmental, community development and policy initiatives. ‘Healthy Schools’ is one of the seven priority areas in the plan.\textsuperscript{112}
Within the plan, the ‘Healthy Schools’ section’s emphasis is on developing movement skills. Another strategy is the revitalisation of school sports programs and improving these programs’ appeal to all students, not just to elite performers. This strategy is written very generally and there is very little detail regarding strategies to increase children’s physical activity engagement and participation during unstructured playtime. There are no specific targets and therefore no mention of evaluation strategies to monitor whether targets have been met.\(^\text{112}\)

Within the NSW Health Centre for Chronic Disease Prevention and Health Advancement, there is a physical activity promotion network with representatives from all Area Health Services in NSW. The network provides a forum for information exchange between its members and supports the creation of evidence regarding the implementation of physical activity promotion projects. NSW Health has also allocated demonstration grants for innovative projects in physical activity promotion since 1996, and some of these grant projects addressed physical activity promotion in schools.\(^\text{116}\)

The author’s experience with one of the NSW Health grants, the Move it Groove it (MIGI) project, was that when schools were faced with a number of physical activity promotion strategies, they tended to choose strategies that focused on individual skill acquisition. The MIGI fundamental movement skills interventions produced significant improvements, while the intervention to increase students’ physical activity levels during PE classes and in the playground during breaks resulted in a small significant change in vigorous PA during PE classes and no significant change in playground PA engagement levels.\(^\text{1, 2, 81, 82, 85}\)

The evidence regarding health promotion interventions, which increased students’ engagement in physical activity during break times, is limited. There is some evidence regarding the positive effects of a teacher-guided structured fitness breaks within an obstacle course framework,\(^\text{117}\) and more evidence regarding the positive effects that playground markings had on students’ physical activity engagement during breaks.\(^\text{118-122}\) The difficulty in getting more evidence is due in part to the complex nature of interventions required to impact on students’ physical activity engagement. Strategies should have a lasting effect on a whole school level, and
address more than one or two cohorts of students. Measuring children’s PA engagement on a whole school level can also be problematic and requires considerable resources and cooperation from schools and the education sector.

**Methodological context**

The preceding review clearly indicates that physical inactivity contributes to obesity and might be amenable to change through health promotion interventions, i.e. policies, environments and behaviours that may mitigate rising population level rates of overweight and obesity. The issue of measuring complex health-related behaviours in relation to such interventions has been widely addressed in the health promotion literature. In the case of physical activity there are a range of methodological approaches. However, once the focus is narrowed to specific contexts such as children’s activity during school breaks, the range of alternatives is considerably reduced.

**Measuring children’s PA engagement during school breaks**

The main focus of this thesis is on measuring children’s physical activity levels in the playground during break or recess playtimes. While research on physical activity levels in school playtime is growing, most studies have small or medium sample sizes. The number of studies that have assessed physical activity levels of large numbers of children in school playgrounds is still small.

Valid and reliable measurement of children’s physical activity levels at the population level cannot be achieved using the most common type of instrument used for adults, the self reported questionnaire. One of the main reasons is that children are less likely to accurately recall physical activity than adults because of developmental differences, especially in abstract thinking and detailed recall. While asking parents to recall their children’s PA engagement might be a viable option for assessment of young children’s physical activity, this method has serious limitations for older children who are not constantly supervised and watched by their parents and definitely cannot be used to assess school time PA engagement.
Consequently, more objective measures have been employed to measure children’s and adolescents’ physical activity levels in contexts such as school playtime. The three main methods are: heart rate monitoring, body motion sensors (accelerometers and pedometers), and direct observation. An excellent discussion of the advantages of, and potential issues with, each of the above measurement methods is provided by Ridgers et al.. The following table summarises the main points of this discussion.

Table 1: A comparison between methods of measuring children’s PA in school playtime

<table>
<thead>
<tr>
<th>Method</th>
<th>Description</th>
<th>Advantages</th>
<th>Potential disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heart rate monitoring</td>
<td>Indication of relative stress placed on the child’s cardio respiratory system.</td>
<td>Cost effective for small sized samples. Relatively unobtrusive.</td>
<td>Difficulty maintaining contact with smaller children’s chest. Affected by non-movement factors, i.e. emotional state, level of fitness.</td>
</tr>
<tr>
<td>Accelerometers</td>
<td>Detect body movement connected with physical activity. Measure and store measurements of the intensity, frequency, pattern, and duration of activity.</td>
<td>Small, unobtrusive, can store large amounts of data to detail activity patterns over time. Valid and reliable</td>
<td>Expensive. High cost for medium to large samples.</td>
</tr>
<tr>
<td>Pedometers</td>
<td>Detect and record steps.</td>
<td>Small, unobtrusive and relatively cheap. Valid and reliable, depending on the brand.</td>
<td>Do not detect movements other than steps. Do not provide information on the pattern or duration of specific activities.</td>
</tr>
<tr>
<td>Direct observation</td>
<td>Assess individual or group physical activity engagement by monitoring frequency, intensity and context of behaviours.</td>
<td>Possible to observe large samples. Contextual factors are included. Valid and reliable, but reliability depends on rigor of training and monitoring. Considered most practical for assessing activity patterns and related factors and behaviours.</td>
<td>May cause reactivity among adolescents. Time consuming. Human resources intensive.</td>
</tr>
</tbody>
</table>
Of the 13 studies found by Ridgers et al., six have used direct observation on individual children and two have used scanning procedures on school playgrounds. Since this review, McKenzie et al. published another study, which uses a whole playground scanning procedure \(^{108}\), and Ridgers et al. published a study, which used heart rate telemetry.\(^{121}\)

**The history and development of CAST**

When the author and his colleagues first started planning the Move it Groove it (MIGI) project in 1997, the promotion of children’s physical activity was discussed mainly at the individual student level. There was an instrument to measure her/his fundamental movement skills mastery.\(^{85, 135}\) There were many instruments to measure an individual’s physical activity level during PE classes.\(^{136, 137}\) These instruments were sometimes used to also measure PA in unstructured playtime of small groups like preschools, but could not do so effectively for bigger groups. Of special relevance to the development of CAST were direct observation instruments, which focused on measuring individual children’s PA. There have been a number of reviews of such instruments and the validation of the physical activity categories they used.\(^{138}\) The author and his colleagues have chosen to use the SOFIT instrument, for the assessment of children’s PA levels in PE classes for the MIGI project, because at the time of planning the MIGI project it had been validated against physiological measures, used in the very large CATCH project, and had incorporated earlier direct observations that McKenzie and his colleagues had previously developed and used.\(^{76, 136, 139}\) The CATCH project resulted in increases in children’s physical activity levels during PE classes\(^{140}\) and improved children’s cardiovascular disease risk profiles.\(^{141}\)

All health promotion practitioners who worked on the project felt that, in accordance with current health promotion theory, MIGI needed to also include strategies that act on a school population level. It was also felt that unstructured break time could provide the biggest opportunity to increase physical activity during the school day.

The project team discussed how to measure children’s engagement in physical activity in the playground. At the time, using motion sensors to measure physical activity engagement of a large number of children was prohibitively expensive. It should be noted that all projects described in this thesis had been implemented and
evaluated by rural health promotion units in NSW. While using accelerometers in bigger epidemiological or intervention studies may be financially feasible, it was not an option for any of the health promotion units. A further discussion of the limitations and practicalities of research in the context of health promotion evaluation is provided in the discussion section.

There were also issues with the measurement bias that may be introduced when children were asked to wear accelerometers or similar motion sensing devices. Since the project was implemented in nine primary schools and had nine control schools, it would have meant transferring the devices from one school to the next. This strategy was deemed too difficult to implement. Heart rate monitors were also deemed to be impractical for studying a large number of children.

Since there were no other published direct observation instruments to measure group level physical activity engagement, it was decided to develop a new instrument for this purpose. There were two main conceptual inputs into this process. The first was provided by Dr Eric van Beurden, the Health Promotion Research and Evaluation Coordinator, who suggested modifying an observational scanning system designed to estimate kangaroo numbers from low-flying aircraft. The author saw potential in combining this system with physical activity classifications from the SOFIT instrument designed to assess individual children’s physical activity level in PE classes and had been validated against heart monitors in a US study. The author and his Australian team subsequently began development of an entirely new system to count groups of children engaged in predetermined levels of physical activity.

This was the impetus for the creation of CAST and for the inclusion of its first validation study in the successful Demonstration Grant proposal submitted to NSW Health in 1998. NSW Health granted the former Northern Rivers Area Health Service Health Promotion Unit almost $200,000 over 2 years to implement the MIGI projects in nine primary schools in Northern NSW. The development of CAST was one of the project’s objectives.
CAST was first trialed by health promotion workers in a school playground in Lismore in 1998. During the MIGI project, the CAST version for five observers was tested and used to evaluate the intervention. Details about the validation and reliability study methodology and results are provided in the respective sections of this thesis and have also been published.1

During the MIGI project, the author had been in contact with Dr Thomas McKenzie who has authored papers about SOFIT and the American CATCH program.83, 105, 106, 136 When CAST was outlined to him, Dr McKenzie informed the author that he and his colleagues were working on a very similar instrument to CAST. In 2000, McKenzie et al. published a paper in which they used SOPLAY to evaluate physical activity engagement levels in schools.124 This paper was the basis for the North Coast team’s intention to drop the number of activity categories scanned from five (lying, sitting, standing, moderately active, and vigorously active) to three (sedentary, moderately active, and vigorously active) following the MIGI project.1

While SOPLAY and CAST were very similar, the SOPLAY instrument was used in middle schools not in primary schools and was used to scan an area on only 2-3 times per visit. The focus of the study was to determine PA engagement in American middle schools, and specifically to compare PA engagement in different venues (eg gyms, basket ball courts) and in different times of the day. It sampled many play areas in a large number of schools in order to derive an estimate of youth engagement in physical activity in middle schools. In contrast, the focus of CAST was to sample many scans in each break so as to produce an estimated physical engagement level for a break period. Another difference was that SOPLAY was only validated against self-reported PA levels while CAST was validated against video analysis of playground physical activity engagement. In later projects CAST was also validated against independent counts of children in the playground as outlined later in this thesis.

Given these differences, SOPLAY provided a very useful support to CAST methodology. However, when health workers from Illawarra Area contacted the author and asked about using CAST in a project, the author suggested to them to use a modified CAST instrument with only three activity categories (sedentary,
moderately active, and vigorously active). While this instrument is similar to SOPLAY, the author felt that SOPLAY’s different sampling and validation method would not have suited the Illawarra study, which sought to estimate mean break period physical activity engagement levels in a small number of schools. The author felt it would be useful to test the reliability and validity of the modified CAST using techniques similar to the ones used in the MIGI project. A training manual for CAST observers from the Central Coast study is attached as appendix 1 and it might be useful to read it before continuing reading the thesis.

The author has also been in contact with project teams from two other areas in which a version of CAST was used. In the Central Coast area of NSW, it was used to evaluate an intervention in six out-of-school care centers. In Dubbo, CAST was used to evaluate an intensive multi-strategic intervention within one school. So far, only baseline data have been collected. Both the Central Coast and Dubbo projects have used an even simpler version of CAST, which has only two activity categories. Since both projects were small and did not have the scope to conduct a video validation study, data were used in this thesis only to test criterion validity in an indirect way and to calculate reliability estimates. As well as these studies where versions of CAST were used, the author is aware of another PhD candidate, who is using CAST for her field research in another project in the Illawarra area.

This thesis is a methodological study, which utilised field projects as an empirical setting. While these projects were evaluated using CAST in quasi experimental designs, the focus of the thesis is not on the evaluation of those projects, but rather on the development, evolution, and testing of an instrument and a number of its versions, to measure physical activity in playgrounds or similar settings. Furthermore, while some instruments are developed in a pure research environment, CAST was iteratively developed in applied settings, using a series of discrete projects, which focused on physical activity in playgrounds.

With this in mind, while project outcomes were more peripheral to the thesis (and are summarised in Appendices 2, 4, and 5), the data and analysis, including data collection, data structure, intra-class correlations, and fitting of multi-level models were key elements of the methodological research. Outcomes for the MIGI and
Illawarra projects have been published elsewhere.\textsuperscript{1, 81, 85, 142} A brief outline of the projects follows, to provide a context of the settings in which the methodological research took place.

**Projects in which CAST has been used**
The broad context of the initial MIGI project has been described in previous sections. In brief, it was conducted by the Health Promotion Unit, of the North Coast Area Health Service, after receiving a grant from NSW Health. It was an intervention study with a quasi-experimental design, involving nine intervention and nine control primary schools. The project's aims were to improve students' fundamental movement skills, increase students' physical activity in PE classes, and increase students' physical activity in school playgrounds during breaks.

The intervention was conducted during 1999 and 2000 and included the establishment of project working groups in each school, and a buddy system with education students assisting teachers regarding both movement skills teaching and increasing physical activity in PE classes. Buddies also participated in the School’s Project Working Group. The Project Working Group’s aim was to work on school-wide approaches to increase physical activity.

Other strategies included teacher training regarding movement skills, dance, and physical activity engagement in PE classes. The project also provided schools with resources such as a web site with scope and sequence plans (including lesson plans and ideas for activities), ideas for increasing overall physical activity levels in the school, and a small grant ($375) to purchase play equipment.\textsuperscript{2} Appendix 6 contains the evaluation report for MIGI.

The Illawarra health promotion workers wanted an instrument that would assist them to measure children’s physical activity engagement during break times in the playground and approached the author regarding using CAST. This was to be used to evaluate an intervention with a small number of schools that redesigned their playgrounds to be safer as well as health promoting. A discussion between Illawarra health promotion workers and the author resulted in the decision to trial a version of CAST with three activity categories and three or four observers (when there was a
fourth observer they counted the overall number of boys or girls in the playground area each scan). Measurements took place in November 2003 and March 2004.

The third project was conducted in the Central Coast in out-of-school-hours child care centers (OOSHCC) by the former Central Coast Area Health Service Health Promotion Unit and NSW Sport and Recreation Department. The project aimed to provide education and resources for child care workers regarding increasing children’s activity levels during their stay in the center after school hours. This study used two or three observers (when there was a third observer they counted the overall number of boys or girls in the playground area each scan).

Pre and post intervention measurements took place in late August/early September and December 2004 respectively in six centres. Due to resource limitations, only two categories of activity were recorded, i.e. whether children were sedentary or active. The project resources did not allow for a validation study, so data from this project were used for reliability estimates only.

The fourth project was conducted by the Greater Western Area Health Service Health Promotion Unit. The aim of the project was to increase students’ physical activity levels in a large primary school in Dubbo, western NSW. The project had a number of more specific objectives, one of which was to increase children’s physical activity engagement in the playground. The project team was especially interested in environmental and design changes to playground areas. A pre intervention data collection was held in November 2005. The intervention was conducted during 2006 and 2007. CAST Follow-up data were not collected following a decision by the Greater Western Area Health Service personnel. As in the Central Coast Project, the Health Promotion Unit conducting the project had limited resources for its evaluation so only reliability estimates are available from data obtained from this project.

Further relevant details from each of the four trial sites will be covered in the Methods section of this thesis. The concluding sections of the introduction review key aspects of reliability and validity in research that underpin the central methodological focus of the thesis.
Measurement reliability and validity in behavioural and social science

Validity
The validity of an instrument is the extent to which the instrument adequately measures or indexes the concepts being studied.\textsuperscript{143, 144} Validity determines whether the research or instrument measures that which it was intended to measure.\textsuperscript{144, 145} Validity can be assessed in a number of ways and “instrument validation is typically cumulative and involves more than one line of evidence in support of the validity of an assessment method”.\textsuperscript{143}

The most detailed discussion of reliability and validity is found in psychological, psychometric and educational literature.\textsuperscript{143} However, numerous instruments have also been developed to measure different dimensions of physical mental, social and spiritual health.\textsuperscript{146, 147} With the rise of the evidence based health care movement, clinicians seek such instruments to rigorously assess health care outcomes. In many cases, results of reliability and validity tests used in the development of the instruments are now available.\textsuperscript{146}

Most instruments, which are mentioned in health care measurement literature, are based on individuals’ perception of aspects of their health. They are mostly administered as written questionnaires.\textsuperscript{146, 147} For example, some of the more widely used instruments in the health promotion field are related to quality of life and social support.\textsuperscript{98} One such instrument is the Short Form Quality Of Life (SF 36) instrument. The SF-36 was developed by the Medical Outcomes Study in the US.\textsuperscript{148, 149} It was developed as an indicator of general health status for use in a population. The instrument has been tested for reliability and validity in Australia.\textsuperscript{150-152}

Other instruments, which are commonly used in health promotion research, are related to health locus of control and self efficacy.\textsuperscript{153-156} Both are related to a person’s perception of their ability to reduce risky behaviours, and adopt and maintain healthy behaviours. The validity of these instruments has been tested in different contexts and for different conditions, eg Schwarzer and Luszcynska (2008) published an overview of validation studies regarding self-efficacy and its
association with nutrition, physical activity and alcohol specific behaviour changes.  

Some types of reliability and validity are easier to measure in the controlled context of a written questionnaire in health care settings, than they are in the context of a direct observation instrument in the field. The following section summarises the concepts of reliability and validity while the Methods section describes how the reliability and validity of CAST were measured with reference to types of reliability and validity covered in the Introduction.

The three main categories of validity are: content, criterion, and construct. Content validity is assessed during an instrument’s early development and, as such, is concerned with conceptual development of the instrument. Trochim (2006) breaks content validity into two types: face and consensual validity and argues that both can be labelled ‘translation validity’ where the focus is conceptual, i.e. researchers think or consult others regarding whether the operationalisation is a good reflection of the construct. Face validity tests whether the instrument ‘looks like’ it measures what it is suppose to measure. Consensual validity relates to the opinion of ‘experts’ regarding whether the instrument meaningfully measures what it is suppose to measure. This form of validity is the subject of literature, which discusses items or constructs incorporated within the instrument.

As mentioned in the ‘History and Development of CAST’ section above, the face validity of CAST was developed by the author and his colleagues as they explored practical and viable options to measure the phenomenon of interest, i.e. children’s PA engagement in school playgrounds.

Criterion validity is the degree of correlation of the measure with another measure of the same phenomenon. Criterion validity can be divided into two types, concurrent or predictive validity. Concurrent validity measures the degree of correlation of two measures of the same phenomenon administered at the same time and predictive validity measures the extent of correlation of a present measure with a future measure. Pennington gives the following example:
“If we find out there is a high correlation between student grades in high-school math classes and their success in college (which can be measured by many possible variables), we would say there is high criterion-related predictive validity between the intermediate variable (grades in high-school math classes) and the ultimate variable (success in college)”. 157

Construct validity involves the empirical and theoretical support for the interpretation of the measured construct or phenomenon. 159 This includes statistical analyses of the internal structure of the test including the relationships between responses to different test items. In health promotion research, construct validity is: “an ongoing effort in which relationships between concepts are tested and revised on the basis of repeated studies”. 143 Statistical analyses of the internal structure of an instrument or test, including the relationships between responses to different test items, are often used to test construct validity, the most common statistical test being the Cronbach alpha coefficient. 160

Other types of construct validity are convergent and discriminant validity. Convergent validity is the correlation of a measure with other measures, which are supposed to measure the same or similar constructs. 158 To establish discriminant validity, one needs to show that measures that should not be related are in reality not related. 158

Reliability
Reliability relates to the quality of measurement, as measured in the consistency, stability and predictability with which an instrument can be administered. 128, 145, 158 The four main types of reliability are: split half reliability, stability over time or test-retest reliability, inter-rater reliability, and internal consistency reliability. 143, 158 144

Split half reliability can be applied to questionnaires, interviews and psychological tests whereby the instrument is split into two halves, which measure the same constructs, eg the Zuckermann scale, in which each half contains a set of questions that both measure sensation seeking tendency. 158 161 Stability over time or test-retest reliability is estimated when the same test is administered to the same sample on two different occasions. 143, 158
Inter-rater reliability is used whenever independent judges or raters are used to quantify some aspect of behaviour.\textsuperscript{158, 162} It is an important type of reliability for this thesis, as CAST was administered by trained raters. Inter-rater reliability measures consistency or agreement between two or more observers or raters who are scoring the same phenomenon. Textbooks often recommend the use of agreement rates to test inter-rater reliability for categorical variables and the Pearson correlation coefficient for testing the reliability of continuous variables.\textsuperscript{143, 158}

Stemler argues that:

“…the general practice of describing inter-rater reliability as a single, unified concept is at best imprecise, and at worst potentially misleading. Rather than representing a single concept, different statistical methods for computing inter-rater reliability can be more accurately classified into one of three categories based upon the underlying goals of analysis. The three general categories … are: 1) consensus estimates, 2) consistency estimates, and 3) measurement estimates”.\textsuperscript{162} (See a more detailed outline of inter-rater reliability issues in the Methods section).

Internal consistency reliability refers to the relationship between items of the same instrument. The internal consistency reliability of an instrument is judged by estimating how well the items that reflect the same construct are correlated and yield similar results.\textsuperscript{143, 158} The most commonly used statistical procedure for testing internal consistency reliability in psychometrics is the Cronbach alpha coefficient.\textsuperscript{160}

While CAST did not lend itself to the classical psychometric definition of internal consistency reliability, the correlation of the two measures of total number of children in the playground (the one obtained by totalling the numbers of children recorded by each observer, and the one obtained by an independent observer in the Illawarra, Central Coast and Dubbo studies) provided another indication of consistency reliability. Since one measure was derived by totalling the number of children recorded by each observer there is some similarity to the concept of different items which comprise a construct. The correlation of the two measures over time in a large number of scans also contributes to the concept of measuring consistency.
Another measure of reliability is the intra class coefficient (ICC) derived from multi-level regression models. It is a measure of correlation, consistency or conformity for a data set when it has multiple groups.\textsuperscript{163,164} When using CAST, the ICC of scans within breaks were used to calculate the reliability of one scan to predict the level of physical activity engagement during a break. Composite reliabilities for predicting the overall break PA level by a given number of scans were also calculated. ICC’s were used as another measure of inter-rater reliability. (See detailed technical outline in the Methods section.)

In light of the forgoing considerations and in order to bound the scope of this research thesis, the primary foci in terms of reliability and validity analysis of the newly developed CAST instrument were consensus, consistency and measurement reliabilities, and face content and criterion validity.

**Summary of the introduction chapter**

The introduction chapter has provided background and context to the research reported in this thesis. It started with a literature review of the epidemiology and health effects of obesity in general and childhood obesity in particular. It then explored the mechanisms involved with obesity, focussed on the environmental factors of food intake, physical activity, sedentary behaviours and the relationship between them. An overview of the development of guidelines for children’s recommended daily physical activity levels was followed by the rationale for promoting physical activity in schools. Physical activity promotion was placed in the broader context of the field of Health Promotion.

The focus then shifted to physical activity related research and in particular research measuring children’s physical activity engagement during school breaks. This led to an outline of the development of CAST, followed by a description of the projects in which CAST has been used. The section was concluded with an overview of the concepts of reliability and validity, details regarding common reliability and validity types and a statement defining the types of reliability and validity addressed in this thesis with respect to the CAST instrument.
The following Methods chapter will detail the CAST field procedures in the four studies. A special emphasis will be given to field procedures related to validation and reliability estimation. It will then describe the methods used in data collection, coding, management and analysis.
2. Methods
This chapter provides details of the CAST field procedures used in the four studies of relevance. Particular emphasis is given to those procedures related to validation and reliability estimation. It will then describe the methods used in data collection, coding, management and analysis, again emphasising those related to validation and reliability estimation of the CAST instrument.

The data referred to in this thesis can be broadly divided into two types: 1. data collected in the field before and after implementation of the different projects to evaluate the intervention effects and determine PA engagement levels, some of which were used to estimate the instrument's reliability, and 2. data collected in the field and by video that were used only to measure the instrument’s reliability and validity.

The opening section of the Methods chapter, describes the administration procedures of the instrument in the field before and after implementation of the different projects. These field/project evaluation data regarding PA engagement were used to calculate the intra class correlation of scans as reliable measures of the full break period. These measures are therefore directly related to the thesis research questions of whether CAST was a reliable and valid instrument to measure break-levels PA engagement of students in school playgrounds.

**Field instrument administration procedures**

The Children activity scanning tool (CAST) was developed to assess PA engagement levels, equipment availability and use, and teacher presence and behaviour during school break times of recess and lunch. The instrument has evolved and has been refined in each additional project in which it was used.

During the MIGI project, CAST was used by a team of five observers to simultaneously and repeatedly scan a play area every 75 seconds (by audio taped signal) till the break ended. For full viewing coverage, each playground was divided into discrete areas which were given equal scanning time. See figures 2 and 3 below for school playground maps, with scanning areas and scan direction. These maps were given to each scanning team at both baseline and follow-up.
observations to ensure that the same playground areas were observed for the same number of scans. Each scan, all observers simultaneously swept the area visually in the same direction. See Appendix 1 for a full description and explanation of procedures.

Figure 2: A sample map of a school playground area with two scanning areas
Figure 3: A sample map of a school playground area with one scanning area (different boundaries were used in recess and lunch according to school rules about different play area boundaries in the different breaks).
All participating schools enforced a ‘no play’ designated time period at the beginning of the lunch break, when students are required to sit down and eat their lunches and are not allowed to play. This rule is monitored by teachers, who are rostered to supervise the playground. In some schools there is a separate bell call for the end of this period and the beginning of the active play period. Observations commenced immediately after the ‘no play’ period finished.

Each observer was asked to monitor one of five PA levels from the System for Observing Fitness Instruction Time (SOFIT) instrument. Briefly, in MIGI these levels include all activities energetically equivalent to 1. lying, 2. sitting, 3. standing, 4. walking, and 5. vigorous physical activity like running and hopping. Scans alternately focused on boys and girls. The task of each observer was to first scan the designated area for the number of children of that gender, engaged in activities classified in the observer’s allocated PA level. In the Illawarra study categories 1-3 were collapsed into one ‘inactive’ category creating an instrument with three levels for three observers: 1. non-active, 2. walking or other moderate activity, and 3. running or vigorous activity. In the Central Coast and Dubbo studies the instrument was further modified to include only two categories: 1. non-active, and 2. active. In the Illawarra, Central Coast and Dubbo studies, in most scans there was also one observer who counted the overall number of boys or girls in the playground on each gender-specific scan. See record sheet in Appendix 1.

Each observer was also allocated an equipment or teacher behaviour category to observe and record by sweeping or scanning the playground visually after they completed their first scan counting children. The observer scanned their assigned equipment availability or usage variable, and after recording the number of equipment items available, or children engaged playing with these items of equipment, the observer might have had to perform another scan to count how many teachers were in the playground and what type of behaviour they were engaged in.

Equipment categories were: number of balls in area; number of equipment items other than balls; number of children playing on fixed equipment (eg, slippery slide, monkey bars are counted as a component each); number of children playing ball
games, i.e. focused on a ball (eye contact and body language/direction) and/or actively manipulating the ball; and number of children playing with non-fixed equipment other than balls.

Teacher behaviour/presence categories were: number of teachers present in area; number of teachers encouraging PA, i.e. verbal encouragement/feedback or teacher’s participation in activities; number of teachers observing including passive supervision; and number of teachers managing including discipline, allocating equipment and active involvement in playground supervision. The teacher behaviour scans tended to be much shorter as there were few teachers in the playground. The number of equipment and teacher variables varied between projects and project teams allocated the equipment categories to observers according to the project research interest and the number of available observers. A training manual for the CAST instrument is available as appendix 1.

Analyses of the relationship between equipment usage and PA engagement in the MIGI project have been published.\textsuperscript{1, 2}

The focus of this thesis is specifically on reliability and validity measures of physical activity engagement. There were insufficient resources to examine reliability and validity of equipment availability & usage and teacher behaviour variables. However, the criterion validity of teacher behaviour categories was assessed against experts’ analysis in the development of the SOFIT instrument.\textsuperscript{136}

In all studies, prior to each scanning session, temperature and relative humidity (by hygrometer/wet bulb thermometer) and ‘wet’ or ‘dry’ day (‘wet’ if evidence of rain prior to or during observation) were recorded along with numbers of available equipment items such as ropes, frisbees, hoops, bats and fixed equipment. Relative humidity and temperature measures were used to calculate Heat Stress Level for each break. This was then included as a variable in all initial regression models to adjust for the possible effect of heat stress on physical activity levels.

Heat stress level measures were not taken during the reliability and validity tests as these data compare the scores of different observers at the same times and
therefore the Heat Stress Level would have been the same for all observers who were scoring. The heat stress variable did not have a direct relationship to reliability and validity measurement and will not be covered further in the thesis.

Sample and settings in which CAST was developed and used

Move It Groove It (MIGI) project – project data
Pilot observations took place in two non-participating schools during 1998. Project field observations took place from late summer to autumn (February to April) in 1999 and 2000. The study population was all children (Years K-6, N=3,912) in 18 Northern Rivers primary schools randomly selected from all schools in the region that had nominated to participate. The final sample of schools was then stratified by school size and Department of Education district.

Illawarra project
Observations took place in November 2003 and March 2004. The study population was all children (Years K-6, N=1,845) in the four primary schools that expressed interest to participate in an intervention that included a safety audit of school playgrounds. The Illawarra Health Promotion Unit staff allocated 2 intervention and 2 control schools.

Central Coast project
Observations took place in winter and summer 2004 (August and November). The study population was all children (Years K-6, N=210) in 6 After Hours Child Care Centres in the NSW Central Coast area. Centres were randomly allocated to intervention or control.

Dubbo project
Baseline observations took place in November 2005. The study population was all children (Years K-6, N=482) in a Primary School in Dubbo. CAST is used as one evaluation method of an intensive physical activity promotion project in this school.
Testers and training

MIGI project
The six CAST observers were all students from the Schools of Education and Exercise Science & Sport Management at Southern Cross University in Lismore, NSW, Australia. The three days training was conducted by the author and his colleague in two stages: 1. identifying SOFIT activity categories and 2. scanning playgrounds using CAST (both video and field experience). Inter-rater consensus measurement regarding SOFIT was repeatedly measured during training until an acceptable agreement rate (>90%) was achieved against an established standard. This involved trainees viewing a videoed PE lesson for which expert observers had previously rated PA levels of selected students. Agreement rates were calculated as the percentage of recorded 20-second observations for which the trainee and an expert observer’s scores were identical. Cohen’s Kappa coefficient was also calculated.

Illawarra project
Five Health Promotion Officers from the Illawarra Area Health Service Population Health Directorate were trained by the author in a similar way to the methods described above regarding the MIGI project. Since the instrument was modified for use in the Illawarra project, another video validation study was conducted by the Illawarra Health Promotion Officers and analysed by the author.

Central Coast project
Seven staff, from the Central Coast Area Health Service and the NSW Department of Sport & Recreation, were trained by the author. Inter-rater reliability testing was conducted by the Central Coast Area Health Service’s Health Promotion Research & Evaluation team.

Dubbo project
A Health Promotion Officer from the South East Sydney & Illawarra Area Health Service trained 3 Health Promotion Officers from the Greater Western Area Health Service following similar procedures to the other projects with guidance and assistance from the author.
Data management and software packages used

**Move It Groove It (MIGI) project**
Research and Evaluation (R&E) staff of the North Coast Area Health Service
Health Promotion Team (NCHP), entered all field and reliability data using the Epi
Info software. Cohen’s Kappa and agreement rate calculations were done in SAS. The author used SAS on
reliability data, to calculate Cronbach alphas for the overall inter-rater reliability score between the raters. SAS was also used to arrange the data in a hierarchical structure. The data were exported from SAS into Microsoft Excel and then exported into SPSS and MLwiN. SPSS was used to further manipulate the data and perform exploratory analyses. MLwiN was used to conduct multi-level regression modelling on the data. The intraclass correlation coefficients (ICCs) from these models were used in this thesis, as were some of the field data results in relation to instrument validity. ICC was calculated using the school and scan variance coefficients from the final regression models using the formula: school level variance/(school level variance + scan level variance). A detailed explanation of ICCs and the formulas used to calculate them is provided later in this chapter.

**Illawarra project**
Project officers from the former Illawarra Health Promotion Unit entered all data collected in this project directly into Microsoft Excel software. The author cleaned the data using Excel and SAS. The author exported the data to MLwiN, conducted multi-level regression analyses of the field project data, and used the intraclass correlation coefficient (ICCs) from these analyses in this thesis. The author used SAS on reliability data, to calculate Cronbach alphas for the overall inter-rater reliability score between the raters. Microsoft Excel was used to calculate Pearson correlation coefficients for data, which compared the overall number of children in the playground as counted by an independent observer with the calculated sum of children counted within each PA category.

**Central Coast project**
Research and Evaluation staff from the former Central Coast Health Promotion Unit entered all data collected in this project into Microsoft Excel. The author cleaned the data using SAS and Excel. The author exported the data to MLwiN, conducted multi-level regression analyses of the field project data and used the
intraclass correlation coefficient (ICCs) from these analyses in this thesis. The author used SAS with reliability data to calculate Cronbach alphas for the overall inter-rater reliability score between the raters. SAS was also used to conduct a principal component factor analysis on the reliability data. ICCs were also used to calculate the similarity between the 7 observers in the Central Coast reliability data using MLwiN. A detailed explanation of formulas is provided later in this chapter. Microsoft Excel was used to calculate Pearson correlation coefficient for data, which compared the overall number of children in the playground as counted by an independent observer with the calculated sum of children counted within each PA category.

**Dubbo project**

Health Promotion staff from the Great Western Area Health Service Health Promotion Unit entered all data collected in this project into Microsoft Excel. The author cleaned the data using SAS and Microsoft Excel. The author used SAS, on reliability data, to calculate Cronbach alphas for the overall inter-rater reliability score between the raters. SAS was also used to conduct a principal component factor analysis on the reliability data. Microsoft Excel was used to calculate Pearson correlation coefficient for data, which compared the overall number of children in the playground as counted by an independent observer with the calculated sum of children counted within each PA category.

Since the focus of this thesis is on the development and different uses of the CAST instrument, the following sections detail first the reliability and then the validity methodology before proceeding to the analysis section.

**Reliability measurements**

**Using a range of methods to measure reliability**

As mentioned in the introduction, a number of reliability measures need to be applied to instruments in order to get an overall picture of their performance. One of the primary foci of this thesis was inter-rater reliability and Stemler (2004) divides relevant measurements into three categories: 1. consensus estimates, 2. consistency estimates, and 3. measurement estimates. These three categories have been applied in different stages of the studies included in this thesis.
Consensus inter-rater reliability regarding activity categories
Inter observer reliability checks for correctly identifying SOFIT categories, were conducted on 28 occasions during the MIGI study. This was done opportunistically as the project included an assessment of PA levels in Physical Education (PE) classes using the SOFIT instrument and the same observers conducted the playground observations using CAST. A check consisted of two observers rating the same child for the same period of time from a common viewing/vantage point. Checks took place at various times of day and in various schools. Data from these reliability checks were analysed using 2 methods: 1. agreement rates for physical activity levels of individual children, and 2. Cohen’s Kappa Coefficient for categorical/dichotomous data. Detailed coverage of the methods and results regarding the Physical Education component of the MIGI project has been published elsewhere (See also Appendix 6).

Unlike in MIGI, where five categories had to be differentiated, in the Illawarra Study, observers only needed to correctly identify whether a child was sedentary, moderately active or vigorously active. In the Central Coast and Dubbo projects, observers only needed to identify whether children were sedentary or active. In the three latter studies, tests regarding identifying PA categories correctly were conducted only during the training program. Once a high agreement rate between all observers was reached, the training progressed to scanning groups of children and comparing these scans for estimating inter-rater reliability. In the MIGI project the CAST observers also participated in another arm of the MIGI study, which used SOFIT to measure individual children’s PA levels during PE classes. Consequently, data measuring consensus inter-rater reliability were available.

Consistency estimates of inter-rater reliability
Since the overall objective of CAST is to estimate the average number of children who engage in physical activity during a break, one way to estimate reliability between raters is to compare the mean number of children recorded by different raters for the whole break. To this end, 95% confidence intervals around the break mean were used to determine whether the break level means were significantly different.
Stemler (2004) argued that “… consistency approaches to estimating inter-rater reliability are more useful when the data are continuous in nature”. 162 During both training (all studies) and field work (Illawarra study) all observers simultaneously scanned and recorded the number of children engaged in one activity category. This was repeated for each category. Pearson Correlation coefficient between paired observers were initially calculated in all studies.

However, Stemler (2004) asserted that using Pearson correlation as a measure of consistency is limited and cumbersome since it has to be done separately for each pair of observers and does not yield an overall estimate. He also argued that the Cronbach alpha Coefficient 160 is a better measure of the overall internal consistency reliability. 162 It is useful for estimating the extent to which scores from a group of raters measure a common dimension or phenomenon. If the Cronbach alpha estimate between raters is low (<0.7), it implies that most of the variance in the composite score is due to error variance and not true score variance.175 Cronbach alphas were therefore calculated in all studies as another and very important indicator of inter-rater reliability.160, 166

In the 3 studies other than MIGI, one observer also counted the total number of children in the playground. This provided another measure of consistency reliability. It was performed by measuring the correlation between the total number of children in the playground area derived by adding the three or two activity categories measured by CAST, and the number of children counted by the independent observer. Pearson Correlation between these two variables was calculated.

It was also a measure of concurrent criterion validity, that is, a comparison of one measure to another measure of the same phenomenon, at the same time. Findings from this analysis will therefore be reported in both the reliability and validity sections of the methods and results chapter.

Another measure of consistency can be derived from data by calculating Intraclass Correlation coefficient (ICC’s) from hierarchical models thereby determining the similarity of scans within schools. ICC’s derived in this manner from the MIGI, Illawarra and Central Coast field data, provided an estimate of the reliability of single
scans as measures of break periods. The formula used to calculate an ICC for a single scan was:

\[ ICC = \frac{\text{school level variance}}{\text{school level variance} + \text{scan level variance}} \].164

To calculate how well the typical number of scans made during a break period predicts the physical activity level of a break, the median number of scans was used in the following formula:

\[ \lambda = \frac{(n \text{ scans} \times \text{single scan ICC})}{1 + ((n \text{ scans} - 1) \times \text{single scan ICC})} \].

For a break of 29 scans (the MIGI median) the ICC formula was:

\[ \lambda_{29} = \frac{(29 \times \text{scan ICC})}{1 + (28 \times \text{single scan ICC})} \].164

As well as calculating ICC’s from field data regarding the reliability of scans in prediction of break PA engagement level, ICC’s were also used to calculate the similarity between the 7 observers in the Central Coast reliability data, following a presentation about the instrument’s reliability at the NSW Physical Activity Network meeting.176 The data were structured hierarchically with seven raters’ scores for each scan, and a variable number of scans in each break. A null regression model was run in MLwiN with PA engagement as the outcome variable and a 2 level structure of raters within scans.169 The variances at the scan and rater levels were used to calculate ICC’s.

While ICC gives the reliability of one rater’s score in predicting the scan’s score, to estimate inter-rater reliability, it would be more useful to measure how well the scores of a number of raters would predict the scan’s score. For example, based on the data for the seven raters, it was possible to calculate how reliably any two raters’ scores predicted the physical activity level of a scan score (derived from the scores of all seven raters of the scan). The formula from the field ICC’s calculation can be used, or another formula, which yielded the same result can be used:

\[ \lambda_2 = \frac{\text{Scan level variance}}{((\text{scan level variance}) + (\text{rater level variance} / 2))} \].164
**Measurement estimates of inter-rater reliability**

“Measurement estimates use the information from all raters (including discrepant scores) when attempting to create a summary score for each unit (like a scan). It takes into account the extent to which each rater influences the score”.\(^{162}\)

A common measurement estimate of inter-rater reliability is computed using principal component factor analysis.\(^{177}\) Using this method, the raters’ scores are used to determine the amount of shared variance in the rating that can be accounted for by the first principal component.

In both the Central Coast and Dubbo studies where principal component factor analysis was used, the first principal component was engagement in PA. The percentage of variance that was explained by the first principal component indicated the extent to which the multiple raters reached agreement. If the shared variance was high then it was likely the raters scored a common construct or phenomenon, eg if the shared variance was 0.8, 80% of the variance between the raters’ scores was due to the scoring of a common phenomenon (PA engagement) and 20% was due to error.

To enable an easy overview of all the reliability measures used in the four studies, they are presented in Table 2.
Table 2: Summary of methods to estimate reliability used in each study

<table>
<thead>
<tr>
<th>Study</th>
<th>Consensus estimates</th>
<th>Consistency estimates</th>
<th>Measurement estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Agreement rate (%)</td>
<td>Kappa Coeff.</td>
<td>95% CI around break mean</td>
</tr>
<tr>
<td>MIIGI</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Illawarra</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Central Coast</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Dubbo</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

1 Simple agreement rate: percentage of observers who agreed on the child’s level of activity (performed in PE classes). Should be >70% to indicate higher reliability level.\(^{162}\)

2 Cohen’s Kappa Statistic.\(^{172}\) Estimates the degree of consensus between raters adjusting for the percentage of agreement that could be expected by chance alone. >0.6 is considered substantial.\(^{178}\)

3 95% confidence intervals around the break mean were used to determine whether the means were significantly different.\(^{173, 174}\)

4 In the Illawarra, Central Coast and Dubbo studies the Pearson correlation coefficient was used for criterion validity and reliability consistency estimates. Typically values >0.7 are acceptable for consistency estimates of inter-rater reliability.\(^{162}\)

5 Cronbach’s alpha allows for an overall estimate of consistency among multiple raters which is suitable for continuous data like the number of active/non-active children. Typically values >0.7 are acceptable for consistency estimates of inter-rater reliability.\(^{162}\)

6 These Intra Class Correlations (ICC’s) were calculated based on the field data in MIIGI, Illawarra and the Central Coast studies. They measured how reliably a single scan of children’s physical activity engagement predicted the engagement level of a break. Composite reliability estimates for the median number of scans in a break were then calculated for engagement in moderate/vigorous activity. The formulas were explained in the consistency measure section above.

7 Only reliability data were used for calculating these ICC’s. They measured how reliably 2 of the 7 testers will predict the physical engagement of a scan as measured by all 7 testers. See explanation and formulas in ICC section above.

8 Principal Components Factor Analysis (PCFA) quantified what percentage of the variance was explained by the first factor (which was the measurement of physical activity engagement in both studies where PCFA was used).
Validity measurements

Activity categories used in CAST were derived from SOFIT and these categories has been confirmed elsewhere. Both types of ‘translational validity’, that is face and content validity, were developed and confirmed in the conceptual and formative research stages.

As outlined in the introduction section, the author and his colleagues considered the face validity of CAST from its conception, and have also searched relevant literature and corresponded with experts in the field of PA measurement of children in schools. Dr van Beurden’s experience with animal abundance studies could be considered as adding to both face validity (being part of the team who developed CAST) and as content validity (adding to the validity of the instrument by citing other research literature). The author’s reading of literature published by McKenzie et al., and consequent contact with Dr McKenzie to specifically discuss children’s PA measurement in school playgrounds, has added to the content validity of CAST.

In a paper that focussed on measuring physical activity, Bauman et al. (2006) state that:

“Validity is a way of describing that the measure is assessing what it is intended to measure. It is often in the form of ‘criterion’ (‘gold standard’) validity, comparing a measure such as a questionnaire to a closer representation of the true underlying phenomenon of interest.”

In the studies covered in this thesis concurrent criterion validity was tested in two ways: 1. a comparison of field scans with a video ‘gold standard’ of scans taken at the same time, and 2. a comparison of total number of children derived from two separate observation processes occurring in the field at the same time, whereby the independent count of number of children was used as the ‘gold standard’.

Video measurement procedures
The criterion validity of the CAST tool was tested in both the MIGI and Illawarra studies as follows:
Field observations were performed in a playground area, which was also videotaped from a convenient vantage point, so each field scan was later compared to an
objective standard score of a video recording, made simultaneously of the same set
of children, in the same playground area. The use of video taping to validate the
field observations was selected as it facilitated a validation of individual physical
activity categories as well as the overall MVPA level. While video taping equipment
was available for the Northern Rivers Health Promotion Unit, funding and time to
obtain a large number of accelerometers were not available. A further discussion of
the limitations and practicalities of research in the context of health promotion
evaluation is provided in the discussion section.

A video standard score was established for percentage Moderate/Vigorous physical
activity during the period observed. This was done by 2 raters scanning the
playground area at given intervals using a counter on the video. The raters counted
the number of children engaged in Moderate/Vigorous physical activity every 5 and
20 seconds in the MIGI and Illawarra studies respectively. If raters’ scores were not
identical, the video was rewound and replayed and raters re-examined the interval
using the pause button or replaying the tape over the same interval a number of
times and focussing on counting children in different sections of the playground.
Raters repeated these actions until they reached consensus regarding how many
children were engaged in Moderate/Vigorous physical activity.

**Analysis procedures**
In the early stages of the video scoring, another pair of raters scored the video
independently to the first pair. After rating 100 scans, the scores of both pairs were
compared using Pearson correlation coefficient. Since the correlation for both MVPA
and VPA were around 0.9, the research team decided that analysis of data for only
one pair was needed to establish the video standard score against which the field
data was compared. See more details in the Results chapter under MIGI inter-rater
reliability.

Pearson correlation and comparisons of mean numbers of children engaged in
MVPA using 95% confidence intervals were used to compare the levels of MVPA
engagement derived from CAST field observations and video analysis. In the
Illawarra project, Cronbach’s alphas were calculated when there were two field
observers whose scores were compared to the video scores (for sedentary and
moderate PA engagement), and Pearson correlation coefficient when only one field
observer’s scores were compared to the video scores (vigorous PA engagement was only scored by one of the observers, as there were only five observers in the Illawarra).

In MIGI, scores from sets of four videoed five-second intervals were averaged to compare to the 20-second intervals used in field observations (the field intervals were identified by the time keeper’s voice on the video, so the video and field intervals could be synchronised). In the Illawarra study both video and field intervals were 20 seconds, so they just had to be synchronised using the timekeeper’s call.

Field-field validation measurement and analysis procedures
Two measures were taken: 1. overall number of boys or girls present in the playground obtained by totalling the numbers of children engaged in each physical activity level and recorded by each observer, and 2. overall number of boys or girls obtained by an independent observer. This was done in the Illawarra, Central Coast and Dubbo studies.

The field-field validation testing assumes that counting the number of children in the playground without needing to determine their level of physical activity engagement is a straightforward task that will yield valid results. While this measure does not validate directly the counting of children engaged in each physical activity category, it does indicate that raters have counted validly by comparing the sum of their separate counts with an independent simultaneous count of the total number of children. Pearson correlation between these two variables was calculated.

Analysis of field data for ICC consistency estimates

MIGI project
Binary variables were computed by recoding the five SOFIT categories to indicate whether a child was engaged in moderate or vigorous physical activity (MVPA and VPA). To accommodate for the similarities of children within scans and scans within schools, a multi-level logistic regression procedure was used via MLwiN software. Schools, ‘scans within schools’ and ‘children within scans’ were treated as random samples and defined as three levels of analysis.
Three models were fitted with MVPA and VPA as the outcome variables: 1. a variance components or null model, 2. a model including only significant predictors, and 3. a model with significant and non-significant variables. The variance components models partitioned the total variance (about the grand mean logit) among the child, scan and school levels, and yielded intraclass correlation coefficients and estimates of measurement reliability. Combinations of independent variables were fitted by stepwise addition to arrive at the significant effects only models, which yielded estimates of the proportions of variance at the school and scan levels that were explained by the set of predictors. Further independent variables, which were considered to potentially account for significant portions of variance, were added to the model to provide tests of their significance.

Illawarra project
Similar procedures to the MIGI project were used. However, since there were only four schools in this study (two intervention and two controls), schools were fitted as dummy variables in a two level model (children within scans). Each dummy school variable was used to create an interaction of a school*pre/post variable to enable comparisons between pre and post changes among the four schools.

Models with significant predictors only were used to generate estimates of the mean percentages and 95% confidence intervals for MVPA and VPA under a number of conditions represented by selected values of the predictor variables. Since schools were entered as dummy variables, the school level variance was explained by the difference between the reference school and the other three schools. As in MIGI the models partitioned the total variance (about the grand mean logit) among the child, scan and school levels (the school level variance was explained by the difference between the reference school and the other three schools since schools were entered as dummy variables), and yielded intraclass correlation coefficients and estimates of measurement reliability.

Central Coast project
Similar procedures to the MIGI project were used. As in MIGI, the data were arranged in a three level structure of child within scan within child-care center. Analysis was performed only regarding the number of active children since the
number of children engaged in vigorous physical activity was not separated from the number of children who were active in this study.

Models with significant predictors only were used to generate estimates of the mean percentages and 95% confidence intervals for MVPA and VPA under a number of conditions represented by selected values of the predictor variables. As in MIGI the models partitioned the total variance (about the grand mean logit) among the child, scan and school levels, and yielded intraclass correlation coefficients and estimates of measurement reliability.

**Dubbo project**

Only baseline reliability data from this study were available to the author within the thesis timeframe. Different types of inter-rater reliability estimates were calculated as described in the inter-rater reliability section above. No further analysis of the field data was conducted by the author.

**Ethics approvals**

Ethics and research approvals relevant to this thesis were:

- Northern Rivers Area Health Service 1999/105 (as part of NSW Health Demonstration Grant application DP 98/1).
- NSW Department of Education & Training SERAP Number: 98-153
- Southern Cross University HREC approval number ECN – 00- 61
- University of Wollongong/Illawarra Area Health Service Human Research Ethics Committee (HE02/040)
- NSW Department of Education & Training SERAP Number: 05-132

**Summary of methods chapter**

This chapter has presented the methods used to explore and answer the research questions. First, the context in which the research took place was detailed through a description of the four studies. This was followed by an outline of the measurement procedures and specifically by detailing how reliability and validity were measured.
The analysis methods and the software packages, which were used in all projects, were presented. The chapter finished with details of the ethics approvals.

The following results chapter will detail the results of the CAST reliability and validity analyses from the four studies. Data structure from these intervention studies will be characterised and intraclass correlation coefficients (ICCs) will be presented as a further measure of consistency reliability of CAST.
3. Results
Sample and settings overview

In total, project field data yielded 54531 child observations made in 2286 scans in 132 break periods in 21 intervention and 20 control primary schools (in the MIGI, Illawarra and Dubbo studies), and six after hours child care centres (in the Central Coast study). For inter-rater reliability testing purposes, 170 scans taken by multiple raters yielded 2383 paired scans (rater compared to each of the other raters). There were 792 scans in the project field data for which there was an independent count of the total number of boys or girls in the playground in the Illawarra, Central Coast and Dubbo studies. In the two validity studies 236 field scans were compared to 788 video scans. School size ranged from 18 to 700 children and child care centres size ranged from 10 to 82 students. In schools, morning recess took place between 10.30am and midday and lunch breaks between 12.35pm and 2.50pm. Observations in childcare centres took place between 3pm and 5pm. Further details regarding the sample and settings in each project are outlined in chapter sections covering individual projects.

Project field data - sample and settings

MIGI project
In total, 17,645 child observations were made in 524 scans in 36 break periods in the 18 schools (total enrolments = 3,912). School size ranged from 18 to 575 children (mean 212). Morning recess took place between 10.30am and midday with a mean observation period of 16 minutes. Lunch break took place between 12.35pm and 2.50pm with a mean observation period of 30 minutes. The gender breakdown of observations was 46.4% female and 53.6% male. The first pilot observations took place in mid 1998. Baseline project field observations were taken during February and March 1999. Follow-up project field observations were taken during November and December 2000.

Illawarra project
In total, 27,904 child observations were made in 1019 scans in 52 recess and lunch break periods in 4 schools (total enrolments = 1,860). School size ranged from 350 to 700 children (mean 460). The gender breakdown of observations was 42.3%
female and 57.7% male. Video-field validity testing was conducted in July 2003. Baseline project field observations were taken during November and December 2003. Follow-up project field observations were taken during March 2004.

**Central Coast project**
In total, 5,602 child observations were made in 631 scans in 6 after school hours day care centres in 36 visits between the hours of 3 and 5pm. Centre size ranged from 10 to 82 students enrolled, but the number of children counted in the playground rarely exceeded 20. The gender breakdown of observations was 50% female and 50% male. Baseline project field observations were taken during August 2003. Follow-up project field observations were taken during December 2003.

**Dubbo project**
In total, 3380 child observations were made in 112 scans in 8 recess and lunch break periods in one school in Dubbo (enrolment = 482). The gender breakdown of observations was 50% female and 50% male. Baseline project field observations were taken during November 2005. No follow-up project field observations were taken.

**CAST reliability - samples and results**

**MIGI project**

*Consensus estimation regarding category identification*
Inter-rater reliability for identifying SOFIT categories was tested. The agreement rate for student activity level ranged from 72.2% to 100% with a mean rate of 96.1%. Cohen’s Kappa coefficient ranged from 0.7 to 1. These results have been published.81

*Consistency estimation regarding number of children in each activity level*
A total of 85 scans were taken by six CAST raters for four activity categories (lying and sitting were combined due to a small number of children engaged in these categories in the playground during the reliability test) yielding 1275 paired scans. There was no significant difference between the mean numbers of children recorded...
by the six raters when 95% confidence intervals around the mean were used. Table 3 summarises these calculations.

**Table 3: Means and 95% confidence intervals (to get lower and upper confidence limits, add or subtract CI from mean) of the six CAST raters’ scores in the MIGI project for four activity categories**

<table>
<thead>
<tr>
<th>Rater</th>
<th>Number of children who were lying or sitting in the playground (20 scans)</th>
<th>Number of children who were standing in the playground (20 scans)</th>
<th>Number of children who were moderately active (20 scans)</th>
<th>Number of children who were vigorously active (25 scans)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Rater 1 2 3 4 5 6</td>
<td>Rater 1 2 3 4 5 6</td>
<td>Rater 1 2 3 4 5 6</td>
<td>Rater 1 2 3 4 5 6</td>
</tr>
<tr>
<td>Mean</td>
<td>6.75 6.4 6.75 6.25 6.35 6.2</td>
<td>11.05 10.38 11.00 9.14 9.52 11.14</td>
<td>5.62 5.86 5.50 6.48 5.19 6.76</td>
<td>3.23 3.15 4.04 3.31 3.46 4</td>
</tr>
<tr>
<td>95% CI</td>
<td>0.50 0.53 0.58 0.46 0.53 0.52</td>
<td>1.86 1.44 2.18 1.49 1.38 1.81</td>
<td>1.69 1.38 1.64 1.28 1.00 1.17</td>
<td>1.07 1.02 1.29 1.16 0.92 1.18</td>
</tr>
</tbody>
</table>

Standardised Cronbach’s alpha coefficients for the number of children who were sitting or lying, standing, moderately active and vigorously active were 0.95, 0.93, 0.93 and 0.94 respectively.

**Inter-rater reliability as part of the video validation study**
A sample of the video intervals (n=100/738 or 13.5%) was analysed by both pairs of video raters independently yielding a 0.91 and 0.89 Pearson correlation for the number of children engaged in MVPA and the percentage of children engaged in VPA respectively.

**Consistency estimation using ICCs – project field data**
The intraclass correlation coefficients (ICC’s measuring the reliabilities of a single scan within a school) were 0.34 for combined moderate and vigorous physical activity (MVPA) and 0.21 for vigorous physical activity (VPA). The median number of scans within schools was 29, yielding school-level composite reliabilities of 0.94 and 0.77 for MVPA and VPA respectively. See also table 11.
Illawarra project

Consistency estimation re number of children in each activity level
Inter-rater reliability was determined based on a total of 19 scans taken simultaneously by five raters yielding 190 paired scans. There was no significant difference between the mean numbers of children recorded by the independent pair of observers when 95% confidence intervals around the mean were used. Table 4 summarises these calculations.

Table 4: Means and 95% confidence intervals of the five CAST raters’ scores in the Illawarra project for three activity categories

<table>
<thead>
<tr>
<th>Number of sedentary children in the playground (seven scans)</th>
<th>Rater</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>24.57</td>
<td>21.00</td>
<td>23.86</td>
<td>23.14</td>
<td>20.57</td>
<td></td>
</tr>
<tr>
<td>95% CI</td>
<td>5.26</td>
<td>6.04</td>
<td>9.53</td>
<td>5.40</td>
<td>6.17</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Number of moderately active children in the playground (four scans)</th>
<th>Rater</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>12.75</td>
<td>15.00</td>
<td>12.50</td>
<td>15.75</td>
<td>17.00</td>
<td></td>
</tr>
<tr>
<td>95% CI</td>
<td>5.39</td>
<td>3.58</td>
<td>3.05</td>
<td>5.02</td>
<td>7.16</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Number of vigorously active children in the playground (eight scans)</th>
<th>Rater</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>5.13</td>
<td>5.63</td>
<td>4.50</td>
<td>5.13</td>
<td>4.75</td>
<td></td>
</tr>
<tr>
<td>95% CI</td>
<td>1.87</td>
<td>0.82</td>
<td>1.66</td>
<td>1.79</td>
<td>1.73</td>
<td></td>
</tr>
</tbody>
</table>

Standardised Cronbach’s alpha coefficients for the number of children who were sedentary, moderately active and vigorously active were 0.97, 0.89, and 0.79 respectively. When observations for all activity categories were combined due to the relatively small number of scans, the standardized Cronbach alpha coefficient was 0.98.

Consistency estimation using ICCs – project field data
The intraclass correlation coefficients (ICC’s measuring the reliabilities of a single scan within a school), were 0.18 for combined moderate and vigorous physical activity and 0.32 for vigorous physical activity (Table 12). The median number of scans within schools was 17, yielding school-level composite reliabilities of 0.79 and 0.89 respectively.
Consistency estimation - total number of boys or girls in the playground

Forty nine scans were undertaken using a fourth observer. The Pearson correlation coefficient between these measures was 0.97.

Central Coast project

Consistency estimation re number of children in each activity level

A total of 40 scans, taken simultaneously by seven observers yielded 840 paired observations of the number of children in a play area (20 scans counting sedentary children and 20 counting active children).

There was no significant difference between either the mean number of active children or the mean number of non-active children recorded by the independent pair of observers when 95% confidence intervals were calculated (Table 5).

Table 5: Means and 95% confidence intervals of the seven CAST raters’ scores in the Central Coast project for two activity categories and the overall number of children

<table>
<thead>
<tr>
<th>Rater</th>
<th>Number of sedentary children in the playground (20 scans)</th>
<th>Number of active children in the playground (20 scans)</th>
<th>An independent count of the number of children in the playground (40 scans)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean 1 2 3 4 5 6 7</td>
<td>Mean 1 2 3 4 5 6 7</td>
<td>Mean 1 2 3 4 5 6 7</td>
</tr>
<tr>
<td></td>
<td>95% CI 1.98 2.30 1.93 2.07 2.30 1.87 2.16</td>
<td>95% CI 2.06 1.66 1.58 2.14 2.05 1.66 1.99</td>
<td>95% CI 2.43 2.31 2.17 2.37 2.35 2.34 2.44</td>
</tr>
</tbody>
</table>

Standardised Cronbach alpha coefficients for the number of sedentary and active children were 0.94 and 0.97 respectively. Mean Pearson correlation coefficient (r) of scan level paired observations was 0.69. The difference between these estimates was due to one rater whose observations returned a much lower r when paired with those of all other raters. See limitations section in the discussion chapter.

Consistency estimation - total number of boys or girls in the playground

During the project’s evaluation, there were 631 field scans for which there was an independent count of the total number of boys or girls in the playground.
Pearson correlation coefficient between the counted number of children in the playground and the number of children, calculated from adding the counts of active and non-active categories, was 0.81.

**Consistency estimation using ICCs – project field data**
The intraclass correlation coefficient (ICC measuring the reliability of a single scan within a centre) was 0.08 for combined moderate and vigorous physical activity (Table 13). The median number of scans within centres was 40, yielding a centre-level composite reliability of 0.63 and 0.77 for the null and final models respectively.

**Consistency estimation using ICCs – reliability data**
The intraclass correlation coefficient (ICC measuring the reliability of a single rater to predict the activity engagement level within a scan as measured by all seven raters), was 0.76 for combined moderate and vigorous physical activity. To estimate inter-rater reliability the author calculated how well the scores of any two observers would predict the mean scan score which was derived by averaging the scan score of all seven raters). This yielded a paired reliability of 0.86.

**Measurement estimation - principal component factor analysis**
The proportion of variance explained by the main factor (i.e. the testers counting the same phenomenon) was 96% for the number of children in the playground, 76% for the number of sedentary children, and 85% for the number of active children.

**Dubbo project**

**Consistency estimation re number of children in each activity level**
Inter-rater reliability was determined based on a total of 26 scans taken simultaneously by 3 CAST observers yielding 78 paired scans.

There was no significant difference between either the mean number of active children or the mean number of non-active children recorded by the independent pair of observers when 95% confidence intervals were calculated. (Table 6)
Table 6: Means and 95% confidence intervals of the three CAST raters’ scores in the Dubbo project for two activity categories and the overall number of children

<table>
<thead>
<tr>
<th>Number of sedentary children (10 scans)</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Rater</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Mean</td>
<td>11.20</td>
<td>9.30</td>
<td>11.50</td>
</tr>
<tr>
<td>95%CI</td>
<td>4.49</td>
<td>4.38</td>
<td>5.17</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Number of active children (10 scans)</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Rater</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Mean</td>
<td>16.60</td>
<td>18.20</td>
<td>16.40</td>
</tr>
<tr>
<td>95%CI</td>
<td>3.29</td>
<td>3.50</td>
<td>4.03</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Number of all children in the playground (six scans)</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Rater</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Mean</td>
<td>33.50</td>
<td>35.17</td>
<td>35.67</td>
</tr>
<tr>
<td>95%CI</td>
<td>4.17</td>
<td>3.66</td>
<td>2.51</td>
</tr>
</tbody>
</table>

Standardised Cronbach alpha coefficient was 0.99 for the number of children in the playground, 0.99 for the number of sedentary children, and 0.98 for the number of active children (moderate and vigorous combined).

There were 112 scans in the project field data for which there was an independent count of the total number of boys or girls in the playground. The Pearson correlation coefficient between the counted number of children in the playground and their number calculated from adding the active and non-active categories was 0.96.

**Measurement estimation using principal component factor analysis**
The proportion of variance explained by the main factor (i.e. the testers counting the same phenomenon) was 98% for the number of children in the playground, 97% for the number of sedentary children, and 96% for the number of active children.

**Summary of the reliability results**

To assist the reader to get an overview of the results of all the reliability measures, they are presented in tables 7 and 8 in the following pages.
### Table 7: Summary of results – reliability estimates used in each study

<table>
<thead>
<tr>
<th>Study</th>
<th>Consensus estimates</th>
<th>Consistency estimates</th>
<th>Measurement estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Agreement rate (%)</td>
<td>Kappa Coef. 2</td>
<td>Pearson coeff. 4</td>
</tr>
<tr>
<td>MIGI</td>
<td>96.1%</td>
<td>0.72-1</td>
<td>Lying &amp; sitting 0.95</td>
</tr>
<tr>
<td>Illawarra</td>
<td></td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>Central Coast</td>
<td></td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>Dubbo</td>
<td></td>
<td></td>
<td>3</td>
</tr>
</tbody>
</table>

1. Simple agreement rate: percentage of observers who agreed on the child’s level of activity (performed in PE classes). Should be >70% to indicate higher reliability level.  

2. Cohen’s Kappa Statistic. Estimates the degree of consensus between raters adjusting for the percentage of agreement that could be expected by chance alone. >0.6 is considered substantial.  

3. 95% confidence intervals around the break mean were used to determine whether the means were significantly different. See table 8 for a full outline of the means and 95% confidence intervals around them for all raters.  

4. In the Illawarra, Central Coast and Dubbo studies the Pearson correlation coefficient was used for criterion validity and consistency estimate whereby independent counts of the number of boys or girls in the playground during a scan, where compared to the sum of the two or three activity categories observed. Typically values >0.7 are acceptable for consistency estimates of reliability.  

5. Cronbach’s alpha allows for an overall estimate of consistency among multiple raters which is suitable for continuous data like the number of active/non-active children. Typically values >0.7 are acceptable for consistency estimates of inter-rater reliability.  

6. These Intra Class Correlations (ICCs) were calculated based on the field data in MIGI, Illawarra and the Central Coast studies. They measured how reliably a single scan of children’s physical activity engagement predicted the engagement level of a break. Composite reliability estimates for the median number of scans in a break were then calculated for engagement in moderate/vigorous activity. The formulas were explained in page 43 in the Methods chapter.  

7. Only reliability data were used for calculating these ICCs. They measured how reliably 2 of the 7 testers will predict the physical engagement of a scan as measured by all 7 testers. See explanation and formulas in the methods section above.  

8. Principal Components Factor Analysis (PCFA) quantified what percentage of the variance was explained by the first factor (which was the measurement of physical activity engagement in both studies where PCFA was used).
Table 8: Mean number of children engaged in physical activity categories and 95% confidence intervals by activity levels in the four studies

Number of sedentary children in the playground by project

<table>
<thead>
<tr>
<th>Project</th>
<th>Number of children who were lying or sitting in the playground (20 scans)</th>
<th>Number of children who were standing in the playground (20 scans)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Move it Groove it study</strong></td>
<td><strong>Rater</strong> 1 2 3 4 5 6</td>
<td>Rater 1 2 3 4 5 6</td>
</tr>
<tr>
<td></td>
<td><strong>Mean</strong> 6.75 6.4 6.75 6.25 6.35 6.2</td>
<td><strong>Mean</strong> 11.05 10.38 11.00 9.14 9.52 11.14</td>
</tr>
<tr>
<td></td>
<td><strong>95% CI</strong> 0.50 0.53 0.58 0.46 0.53 0.52</td>
<td><strong>95% CI</strong> 1.86 1.44 2.18 1.49 1.38 1.81</td>
</tr>
<tr>
<td><strong>Illawarra Study</strong></td>
<td><strong>Number of sedentary children in the playground (seven scans)</strong></td>
<td><strong>Number of moderately active children in the playground (four scans)</strong></td>
</tr>
<tr>
<td></td>
<td><strong>Rater</strong> 1 2 3 4 5</td>
<td><strong>Rater</strong> 1 2 3 4 5</td>
</tr>
<tr>
<td></td>
<td><strong>Mean</strong> 24.57 21.00 23.86 23.14 20.57</td>
<td><strong>Mean</strong> 12.75 15.00 12.50 15.75 17.00</td>
</tr>
<tr>
<td></td>
<td><strong>95% CI</strong> 5.26 6.04 9.53 5.40 6.17</td>
<td><strong>95% CI</strong> 5.39 3.58 3.05 5.02 7.16</td>
</tr>
<tr>
<td><strong>Central Coast study</strong></td>
<td><strong>Number of sedentary children in the playground (20 scans)</strong></td>
<td><strong>Number of moderately active children (10 scans)</strong></td>
</tr>
<tr>
<td></td>
<td><strong>Rater</strong> 1 2 3 4 5 6</td>
<td><strong>Rater</strong> 1 2 3 4 5</td>
</tr>
<tr>
<td></td>
<td><strong>Mean</strong> 9.85 11.7 9.6 10 11.45 9.15 9.6</td>
<td><strong>Mean</strong> 11.20 9.30 11.50</td>
</tr>
<tr>
<td></td>
<td><strong>95% CI</strong> 1.98 2.30 1.93 2.07 2.30 1.87 2.16</td>
<td><strong>95% CI</strong> 4.49 4.38 5.17</td>
</tr>
<tr>
<td><strong>Dubbo study</strong></td>
<td><strong>Number of sedentary children (10 scans)</strong></td>
<td><strong>Number of active children in the playground by project</strong></td>
</tr>
<tr>
<td></td>
<td><strong>Rater</strong> 1 2 3 4 5</td>
<td><strong>Project</strong></td>
</tr>
<tr>
<td></td>
<td><strong>Mean</strong> 11.20 9.30 11.50</td>
<td><strong>Move it Groove it study</strong></td>
</tr>
<tr>
<td></td>
<td><strong>95% CI</strong> 4.49 4.38 5.17</td>
<td><strong>Number of children who were moderately active (20 scans)</strong></td>
</tr>
<tr>
<td></td>
<td><strong>Rater</strong> 1 2 3 4 5 6</td>
<td><strong>Rater</strong> 1 2 3 4 5 6</td>
</tr>
<tr>
<td></td>
<td><strong>Mean</strong> 5.62 5.86 5.50 6.48 5.19 6.76</td>
<td><strong>Mean</strong> 3.23 3.15 4.04 3.31 3.46 4</td>
</tr>
<tr>
<td></td>
<td><strong>95% CI</strong> 1.69 1.38 1.64 1.28 1.00 1.17</td>
<td><strong>95% CI</strong> 1.07 1.02 1.29 1.16 0.92 1.18</td>
</tr>
<tr>
<td></td>
<td><strong>Number of children who were vigorously active (25 scans)</strong></td>
<td><strong>Illawarra Study</strong></td>
</tr>
<tr>
<td></td>
<td><strong>Rater</strong> 1 2 3 4 5 6</td>
<td><strong>Rater</strong> 1 2 3 4 5 6 7</td>
</tr>
<tr>
<td></td>
<td><strong>Mean</strong> 3.23 3.15 4.04 3.31 3.46 4</td>
<td><strong>Mean</strong> 12.75 15.00 12.50 15.75 17.00</td>
</tr>
<tr>
<td></td>
<td><strong>95% CI</strong> 1.07 1.02 1.29 1.16 0.92 1.18</td>
<td><strong>95% CI</strong> 5.39 3.58 3.05 5.02 7.16</td>
</tr>
</tbody>
</table>
### Number of vigorously active children in the playground (eight scans)

<table>
<thead>
<tr>
<th>Rater</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>5.13</td>
<td>5.63</td>
<td>4.50</td>
<td>5.13</td>
<td>4.75</td>
</tr>
<tr>
<td>95% CI</td>
<td>1.87</td>
<td>0.82</td>
<td>1.66</td>
<td>1.79</td>
<td>1.73</td>
</tr>
</tbody>
</table>

### Central Coast study

#### Number of active children in the playground (20 scans)

<table>
<thead>
<tr>
<th>Rater</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>6</td>
<td>5.25</td>
<td>4.7</td>
<td>6.55</td>
<td>5.25</td>
<td>4.85</td>
<td>5.3</td>
</tr>
<tr>
<td>95% CI</td>
<td>2.06</td>
<td>1.66</td>
<td>1.58</td>
<td>2.14</td>
<td>2.05</td>
<td>1.66</td>
<td>1.99</td>
</tr>
</tbody>
</table>

### Dubbo study

#### Number of active children (10 scans)

<table>
<thead>
<tr>
<th>Rater</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>16.60</td>
<td>18.20</td>
<td>16.40</td>
</tr>
<tr>
<td>95% CI</td>
<td>3.29</td>
<td>3.50</td>
<td>4.03</td>
</tr>
</tbody>
</table>

### Overall number of children in the playground

#### Central Coast study (40 scans)

<table>
<thead>
<tr>
<th>Rater</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>16.75</td>
<td>17.05</td>
<td>16</td>
<td>16.7</td>
<td>16.85</td>
<td>16.47</td>
<td>16.6</td>
</tr>
<tr>
<td>95% CI</td>
<td>2.43</td>
<td>2.31</td>
<td>2.17</td>
<td>2.37</td>
<td>2.35</td>
<td>2.34</td>
<td>2.44</td>
</tr>
</tbody>
</table>

#### Dubbo study (six scans)

<table>
<thead>
<tr>
<th>Rater</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>33.50</td>
<td>35.17</td>
<td>35.67</td>
</tr>
<tr>
<td>95% CI</td>
<td>4.17</td>
<td>3.66</td>
<td>2.51</td>
</tr>
</tbody>
</table>

## CAST Validity

### Video-field criterion validity testing - MiGI project

The validation study yielded 186 CAST scans (i.e. one every 20 seconds) and 738 video scans (i.e. one every 5 seconds). Every 4 video scans were averaged to enable comparison with field scans. There was a significant difference between the means for the number of children engaged in MVPA when 95% confidence intervals around the means were calculated. There were no significant differences between video and field means of the number of children counted in the playground and the percentage of children engaged in MVPA when 95% around the means were calculated. See table 9 on the next page.
Table 9: Means and 95% confidence intervals of field and video scores in the MIGI project validation study

<table>
<thead>
<tr>
<th>Number of children engaged in MVPA (186 scans)</th>
<th>Raters</th>
<th>Video</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>8.84</td>
<td>10.16</td>
</tr>
<tr>
<td>95% CI</td>
<td>0.56</td>
<td>0.41</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Number of children (186 scans)</th>
<th>Raters</th>
<th>Video</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>22.38</td>
<td>22.62</td>
</tr>
<tr>
<td>95% CI</td>
<td>0.72</td>
<td>0.62</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Percentage of children engaged in MVPA (186 scans)</th>
<th>Raters</th>
<th>Video</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>39.13%</td>
<td>41.87%</td>
</tr>
<tr>
<td>95% CI</td>
<td>2.11%</td>
<td>1.54%</td>
</tr>
</tbody>
</table>

Pearson correlation coefficient for comparison of field and video scores of number of children engaged in MVPA, total number of children in the playground, and percentage of children engaged in MVPA was 0.47, 0.52 and 0.56 respectively.

Video-field criterion validity testing - Illawarra Project

For criterion validation of the CAST version used in the Illawarra project 50 scans, which yielded 300 paired scans (observer versus video) of the same physical activity category, were assessed. Fifty scans of the overall number of boys or girls in the playground were also assessed. There was no significant difference between the mean number of children engaged in each of the three PA categories and the total number of children in the playground when field observations were compared to video scans and 95% confidence intervals calculated. See following table.
Table 10: Means and 95% confidence intervals of field and video scores in the Illawarra project validation study

<table>
<thead>
<tr>
<th>Number of sedentary children (50 scans)</th>
<th>Rater</th>
<th>1</th>
<th>2</th>
<th>Video</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>4.32</td>
<td>4.64</td>
<td>4.4</td>
<td></td>
</tr>
<tr>
<td>95% CI</td>
<td>0.67</td>
<td>7.0</td>
<td>0.69</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Number of moderately active children (50 scans)</th>
<th>Rater</th>
<th>1</th>
<th>2</th>
<th>Video</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>2.19</td>
<td>2.06</td>
<td>2.62</td>
<td></td>
</tr>
<tr>
<td>95% CI</td>
<td>0.49</td>
<td>0.43</td>
<td>0.48</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Number of vigorously active children (50 scans)</th>
<th>Rater</th>
<th>Video</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>2.19</td>
<td>2.62</td>
</tr>
<tr>
<td>95% CI</td>
<td>0.69</td>
<td>0.68</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Number of children in playground (50 scans)</th>
<th>Rater</th>
<th>Video</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>8.76</td>
<td>8.74</td>
</tr>
<tr>
<td>95% CI</td>
<td>0.69</td>
<td>0.68</td>
</tr>
</tbody>
</table>

Standardised Cronbach’s alpha coefficients for comparison of video scores and two independent field observers were 0.98 and 0.85 for numbers of sedentary and moderately active children respectively. Pearson correlation coefficient for comparison of video scores and one field observer of engagement in vigorous physical activity was 0.93. Pearson correlation coefficient for comparison of video scores and one field observer, of total number of children in the playground, was 0.89.

Field-field criterion validity testing
As outlined in the methods section, in three of the four studies, some or all field scans had also an extra observer who independently counted the number of boys or girls present in the playground area without regard to their physical activity engagement. A Pearson correlation coefficient was calculated in those scans between the total number of children calculated by adding the number of children in each of the PA categories observed, and the total number of children counted by the independent observer. These results have been reported in the reliability section as well.
**Illawarra project**
Forty nine scans of the field scans were undertaken using a fourth observer. The Pearson correlation coefficient between the total number of children calculated by adding the number of children in each of the three PA categories and the total number of children counted by the fourth observer was 0.97.

**Central Coast**
There were 631 field scans for which there was an independent count of the total number of boys or girls in the playground. The Pearson Correlation Coefficient between the counted number of children in the playground and the number of children, calculated from adding the counts of active and non-active categories, was 0.81.

**Dubbo Project**
There were 112 scans in the project field data for which there was an independent count of the total number of boys or girls in the playground. The Pearson correlation coefficient between the counted number of children in the playground and their number calculated from adding the active and non-active categories was 0.96.

**Physical activity engagement field data, which were relevant to CAST reliability**

In the following section, relevant results from three projects are presented. Since the scope of this thesis is on measurement issues of CAST, only the variables' estimates are presented. Of particular relevance are the intraclass correlation coefficient (ICC) estimates of scans, as they indicate the reliability of scans as measures of the mean PA engagement during a whole break. This section was positioned at the end of the results chapter after the main reliability and validity results were presented. If the reader is interested in the findings of the field studies in terms of intervention effects and other significant predictors, the MIGI findings are presented in Appendices 2 and 6, the Illawarra findings are presented in Appendix 4, and the Central Coast findings are presented in Appendix 5. No post intervention data from the Dubbo study were available.
**MIGI project**
Results of models are reported in table 11. Parameter estimates are considered to be significant if they are at least 1.96 or 1.64 times their standard errors (two-and-one tailed respectively at p=.05). The significant fixed effects predictors are marked with *.

*Table 11: Parameter estimates and standard errors from variance components and significant effects only for MVPA and VPA in the MIGI project*

<table>
<thead>
<tr>
<th>Fixed effects</th>
<th>MVPA Null Model¹</th>
<th>MVPA Predictor Model²</th>
<th>VPA Null Model¹</th>
<th>VPA Predictor Model²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-0.227 0.099</td>
<td>-0.360 0.491</td>
<td>-2.145 0.108</td>
<td>-1.316 0.596</td>
</tr>
<tr>
<td>Recess³</td>
<td>-0.149* 0.076</td>
<td>-0.296* 0.097</td>
<td>-0.552* 0.081</td>
<td>-0.164* 0.063</td>
</tr>
<tr>
<td>Girl child⁴</td>
<td>-0.413* 0.062</td>
<td>-0.552* 0.081</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Enrolled⁵</td>
<td>-0.121* 0.053</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Balls⁶</td>
<td>-0.121* 0.053</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other equip.⁶</td>
<td>0.000 0.005</td>
<td>0.003 0.004</td>
<td>0.003 0.004</td>
<td>0.000 0.004</td>
</tr>
<tr>
<td>Fixed equip.⁶</td>
<td>0.003 0.003</td>
<td>0.000 0.004</td>
<td>0.000 0.004</td>
<td>0.000 0.004</td>
</tr>
<tr>
<td>Encourage⁷</td>
<td>-0.008 0.034</td>
<td>-0.025 0.044</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observe⁷</td>
<td>0.033 0.021</td>
<td>0.044 0.026</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Manage⁷</td>
<td>0.009 0.028</td>
<td>-0.019 0.036</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Heat stress⁸</td>
<td>0.005 0.013</td>
<td>-0.015 0.016</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Random effects ⁹</th>
<th>Coeff.</th>
<th>S.E.</th>
<th>Coeff.</th>
<th>S.E.</th>
<th>Coeff.</th>
<th>S.E.</th>
<th>Coeff.</th>
<th>S.E.</th>
</tr>
</thead>
<tbody>
<tr>
<td>School</td>
<td>0.158  0.059</td>
<td>0.080  0.034</td>
<td>0.179  0.070</td>
<td>0.112  0.049</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scan</td>
<td>0.301  0.030</td>
<td>0.230  0.027</td>
<td>0.413  0.051</td>
<td>0.273  0.043</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Child ¹⁰</td>
<td>1.000  0.000</td>
<td>1.000  0.000</td>
<td>1.000  0.000</td>
<td>1.000  0.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Statistics:**

\[
\begin{align*}
\text{ICC} &= r_1^{11} = 0.344 \\
r_2^{12} &= 0.938 \\
R_1^{213} &= 0.325 \\
R_2^{214} &= 0.478 \\
\end{align*}
\]

¹ Variance components model

² Model with significant and non significant variables
3 Break period (reference category = lunch)
4 Dummy code to represent child gender (reference category = male)
5 School size - number of children enrolled (school deviation from mean = 212)
6 Variables measuring equipment availability (items per 100 children)
7 Variables measuring teacher behaviour/presence (teacher behaviours per 100 children)
8 Variable measuring heat stress
9 Variance components and their standard errors (SE)
10 These values are default, given that observations on children within scans provide scan logits
11 ICC = intraclass correlation coefficient = \( r_1 \) = reliability of a single scan as measure of a school
12 \( r_2 \) = reliability of mean of 29 (the median) scans as measure of a school
13 \( R_1^2 \) = proportion of scan variance accounted for
14 \( R_2^2 \) = proportion of school variance accounted for

The null or variance components models reveal considerably more variance at the 'scan within school' than at the 'school' level (i.e., there is more variation within than between schools), with 34% of the variance of MVPA and 21% of the variance of VPA located at the 'scans within schools' level. Since the between schools variance is a variance of the schools’ mean engagement levels, it is plausible that the scans within each school will be more varied than the school means.

**Illawarra project**
Results of models are reported in table 12. Parameter estimates are considered to be significant if they are at least 1.96 or 1.64 times their standard errors (two-and-one tailed respectively at \( p = .05 \)). The significant fixed effects predictors are marked with *. 
Table 12: Parameter estimates and standard errors from variance components and significant effects only for MVPA and VPA in the Illawarra project

<table>
<thead>
<tr>
<th>Fixed effects</th>
<th>MPA Null Model¹</th>
<th>MPA Predictor Model²</th>
<th>VPA Null Model¹</th>
<th>VPA Predictor Model²</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coeff.</td>
<td>S.E.</td>
<td>Coeff.</td>
<td>S.E.</td>
</tr>
<tr>
<td>Intercep b₁j</td>
<td>-0.585 0.020</td>
<td>-0.821 0.043</td>
<td>-1.749 0.029</td>
<td>-1.657 0.113</td>
</tr>
<tr>
<td>Prepost³ b₂j</td>
<td>0.253* 0.060</td>
<td>-0.646* 0.083</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender⁴ b₃ij</td>
<td>-0.189* 0.038</td>
<td>-0.410* 0.054</td>
<td></td>
<td></td>
</tr>
<tr>
<td>School2Dum⁵ b₄j</td>
<td>0.531* 0.069</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>School3Dum⁵ b₅j</td>
<td>0.600* 0.068</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>School4Dum⁵ b₆j</td>
<td></td>
<td></td>
<td>-0.663* 0.102</td>
<td></td>
</tr>
<tr>
<td>Prepost*Sch² b₇j</td>
<td>-0.257* 0.098</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prepost*Sch³ b₈j</td>
<td>-0.609* 0.100</td>
<td>1.291* 0.142</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prepost*Sch⁴ b₉j</td>
<td>0.169* 0.070</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Girl<em>PP</em>Sch² b₁₀ij</td>
<td></td>
<td></td>
<td>0.810* 0.168</td>
<td></td>
</tr>
<tr>
<td>Girl<em>PP</em>Sch³ b₁₁ij</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Girl<em>PP</em>Sch⁴ b₁₂ij</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Break/Recess 8 b₁₃j</td>
<td></td>
<td></td>
<td>0.199* 0.059</td>
<td></td>
</tr>
<tr>
<td>ProplayBall 9 b₁₄ij</td>
<td></td>
<td></td>
<td>0.550* 0.220</td>
<td></td>
</tr>
<tr>
<td>ProplayEquip 9 b₁₄ij</td>
<td></td>
<td></td>
<td>0.166* 0.024</td>
<td></td>
</tr>
<tr>
<td>Adj Heat stress¹⁰ b₁₅j</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scan u₁j</td>
<td>0.221 0.018</td>
<td>0.155 0.015</td>
<td>0.459 0.035</td>
<td>0.305 0.028</td>
</tr>
<tr>
<td>Child 12 e₀ij</td>
<td>1.000 0.000</td>
<td>1.000 0.000</td>
<td>1.000 0.000</td>
<td>1.000 0.000</td>
</tr>
<tr>
<td>Statistics:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ICC=r₁¹³</td>
<td>0.181</td>
<td>0.315</td>
<td></td>
<td></td>
</tr>
<tr>
<td>r²¹⁴</td>
<td>0.790</td>
<td>0.886</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R²¹⁵</td>
<td>0.054</td>
<td>0.106</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

¹ Variance components model
² Model with significant and non significant variables. Only significant variables and interactions estimates are included.
³ Pre or post intervention measurement (reference category = pre)
⁴ Dummy code to represent child gender (reference category = male)
⁵ Dummy code for schools 2-4 (reference category = school number 1, schools 1&2 intervention, Schools 3&4 control)
6 Dummy code for interaction variables between Pre/Post and school number (reference category=school number 1 at pre)
7 Dummy code for interaction variables between gender, Pre/Post and school number 4 (reference category = boys Pre intervention in school number 1}
8 Variable indicating break type (recess or lunch) (reference category = recess) heat stress
9 Variables measuring the number of children playing with balls or other equipment (eg ropes, frisbees, bats, but excluding fixed equipment) as a proportion of total number of children scanned
10 Adjusted Heat Stress = Heat Stress level deviation from the mean (Mean HSL = 76.6). Using adjusted Heat Stress level instead of Heat Stress level makes it much easier to understand the models' results.
11 Variance components and their standard errors (SE)
12 These values are default, given that observations on children within scans provide scan logits
13 ICC= Intra Class Correlation = r1 = reliability of a single scan as a measure of a school
14 r1 = reliability of mean of 17 scans (the median number of scans per break) as a measure of a school
15 R2 = proportion of scan variance accounted for by the model

Central Coast project
Results of models are reported in table 13. Parameter estimates are considered to be significant if they are at least 1.96 or 1.64 times their standard errors (two-and-one tailed respectively at p=.05). The significant fixed effects predictors are marked with *.
Table 13: Parameter estimates and standard errors from variance component and significant effects only models for MVPA

<table>
<thead>
<tr>
<th></th>
<th>Model 1 (variance components only)</th>
<th>Model 2 (significant variables only)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coefficient</td>
<td>Std Error</td>
</tr>
<tr>
<td><strong>Fixed Effects</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_{1jk} = \text{Intercept}$</td>
<td>-0.336</td>
<td>0.063</td>
</tr>
<tr>
<td>$\beta_{2jk} = % \text{children playing ball games}$</td>
<td>1.500*</td>
<td>0.148</td>
</tr>
<tr>
<td>$\beta_{3jk} = % \text{children playing with fixed equipment}$</td>
<td>0.512*</td>
<td>0.127</td>
</tr>
<tr>
<td>$\beta_{4jk} = n \text{workers present in playground}$</td>
<td>-0.116*</td>
<td>0.049</td>
</tr>
<tr>
<td><strong>Random Effects</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$v_{1k} = \text{centre}$</td>
<td>0.015</td>
<td>0.014</td>
</tr>
<tr>
<td>$u_{2jk} = \text{scan}$</td>
<td>0.353</td>
<td>0.049</td>
</tr>
<tr>
<td>$e_{0jk} = \text{child}^1$</td>
<td>1.000</td>
<td>0.000</td>
</tr>
<tr>
<td><strong>Statistics</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\text{ICC} = r_1^2$</td>
<td>0.041</td>
<td></td>
</tr>
<tr>
<td>$r_2^3$</td>
<td>0.630</td>
<td></td>
</tr>
<tr>
<td>$R_{12}^4$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$R_{22}^5$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

1 These values are default given that observations on children within scans provide scan level logits
2 ICC = intraclass correlation coefficient = $r_1$ = reliability of a single scan as measure of a centre
3 $r_2$ = reliability of mean of 40 (median number of scans per break) scans as measure of a centre
4 $R_{12}^2$ = proportion of scan variance accounted for
5 $R_{22}^2$ = proportion of centre variance accounted for

**Dubbo project**

Only baseline reliability data from this study were available to the author within the thesis timeframe. Different types of inter-rater reliability estimates were calculated as described in the inter-rater reliability section above. No further analyses of the field data were conducted by the author.

**Summary of results chapter**

This chapter has presented the findings of all analyses conducted in order to address the research questions posed in the introduction. Specific analyses were used to tests the associated hypotheses. Part of the methodological analysis involved comparison between versions of CAST using a different number of observers and measuring a different number of physical activity categories.
This involved initial assessment of reliability and validity of its usage with multiple observers and physical activity categories. Subsequently, data from two other smaller studies were used to assess CAST reliability, when performed by only two observers and measuring only two physical activity categories (active and sedentary).

In the final Discussion chapter the main research questions and hypotheses will be revisited in the light of the foregoing analyses and their findings.
4. Discussion
This, the final chapter of the thesis, will revisit the main research questions and hypotheses in the light of the foregoing analyses and their findings. It will address the strengths and weaknesses of the overall study and its component parts and will link the main findings back to pertinent points raised by other researchers in the field. It will make recommendations relating to current measurement issues, intervention and evaluation strategies and relevant policies and guidelines relating to child physical activity in the specific context of school playgrounds. It will also highlight issues that the author feels require the further attention of researchers in the field.

**Answering the research questions**

The findings outlined in this thesis lead to rejection of the null hypotheses and therefore confirm the main research questions. Physical activity engagement levels of children in school playgrounds and similar environments could be reliably and validly measured using a direct observation instrument. More specifically, the CAST instrument was a valid instrument to measure children's physical activity levels in school playgrounds, and its reliability was confirmed using a number of reliability measures in four studies.

Criterion validity against a video was only tested on the five and three categories versions of CAST. Criterion validity of the two categories version was tested against an independent count of children in the playground. Reliability estimates were high for all versions. Further details, regarding the features of the different versions and a summary of the findings are presented in the features section below and table 14. Further details, regarding the limitations of both validity and reliability findings, are presented in the limitations section.
CAST’s relevance to emerging evidence - PA during school breaks and other uses

The introduction chapter outlined the importance of increasing physical activity for general health and child development generally, and arresting the childhood obesity epidemic particularly. In a large cohort study, Ness et al. (2007) found significant inverse association between children's physical activity engagement and obesity in both genders. As already presented in the introduction chapter, the trends of children’s physical activity levels since the 1980s differ across countries, but significant upward trends of childhood and youth overweight and obesity are present in both developed and developing nations.

Since the time spent in sedentary pursuits, especially watching TV, has been found to be correlated with increased risk of being overweight and obese among children, increasing children’s daily physical activity levels is a crucial strategy to the success of childhood obesity prevention programs.

In this context, promoting children’s PA during school hours is an important obesity prevention objective. Ridgers et al. (2006) estimated that PA during recess times may contribute up to half of children’s daily PA requirement (which was one hour at the time of her paper’s publication, but may be revised upwards, if Andersen et al.'s recommendations were adopted). Ridgers et al. concluded that it would be a realistic target for schools to achieve the engagement of children in MVPA for 40% of recess times in schools.

If such recommendations were to be adopted by education authorities, health promotion practitioners, teachers and parents, a practical and accurate method to measure PA engagement in school playgrounds would be required. CAST could be used for this purpose.

Due to the relative newness of the field of physical activity measurement in general, and measurement of children’s PA in schools in particular, there is no consensus regarding the best measurement methods. Both Miller (2004) and Welk (2000) refer to the factors influencing decisions regarding which instrument to use, as the
Advantages and disadvantages of heart rate monitors, pedometers, accelerometers and direct observation instruments have been summarised in table 1 in the introduction chapter.

It is the author’s opinion that the decision regarding which instrument to use, to measure children’s PA levels in schools, would depend on the research questions, the resources available to the researchers, and the context in which the research is conducted. It is also the author’s opinion that CAST and other direct observation instruments may offer a practical and low cost measurement solution for schools or educational authorities.

While pedometers, and to a lesser degree accelerometers, may provide additional motivation for study participants to increase their PA levels, both have limitations. Pedometers have reliability and validity issues due to their inability to distinguish between different types of movements. Reliability and validity of accelerometers has improved recently and they are the preferred measurement tool for recent epidemiological studies. However, they cannot accurately record an activity which does not involve the whole body, e.g. bouncing a ball while standing, while observational studies can. Another important limitation of the accelerometers is their cost. The ActiGraph accelerometer used in Ness et al.’s study was advertised on ActiGraph web site for over US$300 at the time of writing. This would be prohibitively expensive as a means of measurement in a health promotion project if activity levels of a large number of children were to be measured in the playground. Very well funded research projects could use accelerometers to validate direct observation instruments like CAST.

Health promotion or education staff can be trained to use CAST reliably. However, as the Central Coast study illustrated, there is a need for thorough training and continuous monitoring of reliability data. The author also suggests at the end of this chapter that other even simpler versions of CAST should be tested.

The fact that there has been a demand for using CAST in a number of different projects points to its potential role in creating further evidence regarding PA engagement in school playgrounds or similar environments. The author is aware of
four projects, other than MIGI, that have used CAST in NSW (three of which were outlined in this thesis and another project conducted by a PhD candidate from Wollongong University). The author has also provided the CAST training manual and related resources to researchers from the UK, Greece, Belgium, and Hong Kong.

Health promotion workers in the North Coast Area Health Service approached the author with a query regarding using CAST to evaluate a physical activity promotion school grant project called Grant of School Health (GOSH). The project plan included awarding grants to schools to implement a PA promotion project within the school. It also required that students identify barriers to PA engagement in their school, and that schools respond by ameliorating or removing these barriers. The author’s colleagues suggested that either teachers or year 6 students (11-12 years of age) could be trained to use CAST so that pre and post intervention measurements could be undertaken to evaluate the program’s effectiveness. Just before the submission of this thesis, a health promotion worker, from the Illawarra Area Health Service, approached the author regarding training teachers to use CAST in a new project.

If this were to eventuate then the CAST based evaluation would become part of an intervention whereby schools would use action learning and organisational change principles to carry out the implementation. There would be a process of identifying barriers to PA, planning and taking action to address these barriers, evaluating the effect of this action, and then starting a new cycle of planning and implementing action.

Such a project design has a strong appeal to health promotion practitioners, as it is more likely to produce sustainable changes that will increase PA engagement in schools. Using CAST in such contexts will provide feedback to teachers, parents and students regarding the success or otherwise of their programs regarding changing PA engagement levels. Likely changes would occur at the organizational level as well as through environmental modification. Using CAST as part of the process could also educate teachers and students about PA engagement levels of children in their school.
A disadvantage of using students to use CAST during break times is that by doing so they will be reducing their own physical activity engagement during these periods. Careful consideration to scheduling of children observers will need to be practiced. It might be possible to test the effect that being children’s CAST observer would have on their PA engagement. The balance between the (hopefully positive) long term effect of being a CAST observer on children’s PA engagement and the reduction of their PA engagement while observing playgrounds will have to be clearly in favour of improved PA engagement following participation as an observer, for the project to be ethically sound.

The author has cautioned his colleagues that using CAST rigorously will require thorough training and continuous monitoring of reliability data, at least with the first cohort of teacher or student observers. Such monitoring is important to counter potential rater bias stemming from observing one’s own school, as well as the process of observers’ drift, in which raters ‘drift’ after a period of using the same instrument and either over or under estimate. Such drift may be stronger when observing one’s students and/or peers due to close first-hand knowledge of, and empathy with their behavioural characteristics. There are, however, no data in this study to illuminate the extent of this risk.

As mentioned in the results section regarding the Central Coast data, calculating Cronbach alpha for the correlation of individual raters’ counts, with the mean count of seven raters, revealed that one rater was recording significantly differently to the other six. This had an influence of the mean $r$ of all paired correlations, which was much lower than the Cronbach alpha estimate (0.69 and 0.93 respectively). This made sense as Cronbach alpha measured the overall inter-rater reliability, which was generally very high when all seven raters were considered. The Pearson correlation on the other hand was strongly influenced by the one rater whose scores did not correlate with all other scores.

If students were to use CAST, such testing would be essential. This will necessitate an initial separate research process to determine whether senior primary school
students are capable of using CAST reliably, but would also require further studies to monitor raters’ reliability.

A decision whether to embark on a school based project, in which CAST is used to evaluate playground PA engagement during play time, will need to allow for formative and continuous research into children’s ability to use the instrument in the first year of two of the project. Another factor, which needs to be considered, is the consistent use of raters for pre and post observations. In all projects described in this thesis an effort was made to use the same team of observers for pre and post observations. However, an exact match was not possible on all occasions for logistical reasons. To increase the similarity of pre and post observations, the playground areas observed and direction of scanning for lunch and recess were clearly marked on all school maps (see figures 1 and 2). The number of scans per area were also clearly recorded either on school maps or related school data forms.

Since even the best inter-rater reliability results were not a perfect 1, it is recommended that future studies that use CAST use the same teams of observers for pre and post observations as much as possible to minimize rater-specific bias. However, the inter-rater reliability results for all studies were high, so findings will be reliable even when some of the observations are performed by teams with different composition of raters and pre and post observations.

An advantage of this approach is that involving students or teachers in measuring PA might in itself lead to increased PA engagement, as it will give them insight into PA engagement patterns and ways to change them. A purely positivist approach to research may view such involvement of researchers in the subject under investigation as inappropriate\textsuperscript{186}. However, taking an action learning approach to such a school project would naturally use CAST as an evaluation tool, as well as an implementation strategy to increase key informants’ insight into issues related to playground PA engagement.\textsuperscript{184} Using CAST to observe two activity categories (sedentary, active) by teachers or students will be less demanding on school resources than using the three-category version, especially if teachers were to perform the observations (a version of CAST in which one observer records the two categories consequently could be investigated, see recommendations). However, if
schools were interested to distinguish between moderate and vigorous physical activity engagement the three categories version would be required, placing a greater demand on school resources.

Further research and development of CAST, as detailed later in this chapter, could potentially result in a more affordable, easier to use, valid and reliable instrument to measure PA in school playgrounds and similar environments.

**Contribution to the PA measurement field and recommendations for future uses**

While CAST has been used primarily to measure physical activity engagement in school playgrounds, it has potential uses in other settings both within and outside schools.

Within schools, CAST can be used to assess physical activity engagement levels in more structured environments than school playgrounds such as physical education (PE) classes. This might provide data regarding the different engagement levels following certain teaching strategies or class activities.

There has been a strong focus, in recent years in NSW, on improving children’s fundamental movement skills in PE classes. Improving movement skills is a very important way to increase future participation in organised sport and other physical activity engagement. However, teachers need to strike a balance between teaching fundamental movement skills and providing opportunities for children to be active during classes.

While there are already valid instruments to measure children’s physical activity levels during PE classes, they mostly rely on observations of individual children. Such methods may not give an overall picture of the class’s physical activity engagement level. Observing individual children may be appropriate for evaluation of an intervention, which occurs in a number of classes. The sampling of a number of children per class gives an overall mean level of physical activity in physical education classes. It does not necessarily give an accurate estimate of the activity
level in a single class. Using CAST may provide a better overall assessment of PA level in individual classes.

Another advantage of using CAST would be that if teachers were trained to use it they would start to get a ‘feel’ regarding physical activity engagement levels. They would therefore be more aware of the impact that different activities and educational strategies have on physical activity engagement within their class.

CAST can be useful to gauge children’s physical activity engagement when participating in different sporting activities and games. Such awareness can assist teachers and coaches to choose games and codes that typically lead to engagement of a higher proportion of participants, or modify existing game rules to encourage higher engagement levels, for example through smaller team size.

The most likely version of CAST to be used in schools settings would be the two observers version, which requires fewer resources. A version using only one observer, if it was tested as per the recommendations further in this chapter, may also be preferable for schools. However, if CAST was used in the context of school sporting activities and games by coaches who may be interested in building children’s cardiovascular fitness, the three categories version may be required to provide data regarding engagement in vigorous physical activity.

Outside schools, CAST can be used to assess physical activity engagement levels in recreational facilities and settings such as parks, public playgrounds and sport fields. Feedback regarding the public’s engagement in physical activity in such settings may be important for planning and policy purposes, to bodies such as local governments and sport and recreation departments. Using CAST in such surroundings would be most rigorous if some validity and reliability tests specific to such environments were conducted. To maximise reliability and validity, similar processes to the ones described in the methods section would be useful, i.e. having clear definition of boundaries of area observed, direction of scanning and method of signalling commencement of scanning. Another aspect of using CAST outside schools, which will require testing, is reactivity of subjects. Sallis and Owen (1999) argue that direct observation instruments may elicit reactivity and may even be
unacceptable to adolescents and adult subjects.\textsuperscript{134} However, they do not specify the number of observers or type of observation. The author believes that it would be useful to test the acceptability of CAST or similar instruments, when conducted by one or two people, in recreational settings and facilities.

CAST can obviously be used to evaluate physical activity promotion programs when the program objective is to increase physical engagement of a group in unstructured settings as has been demonstrated in the studies presented in this thesis. As mentioned in the previous paragraph, using CAST to evaluate interventions can extend beyond school playgrounds to other settings. In the context of program evaluation CAST can be used in both longitudinal and cross-sectional studies depending on the unit of analysis. CAST can be used to compare the same classes or schools over a number of points in time, but cannot track individual children. Since the focus of interest in population health project evaluation is often on the class, school or population mean, it can be used effectively to track changes at these levels over time.

\textbf{Contribution to playgrounds direct observation PA measures - comparison to other instruments}

This thesis contributes to the literature about direct observation instruments to measure PA engagement at the level of large groups. The validity of instruments which measure children one at a time has been established via comparison to physiological based instruments like heart rate monitors.\textsuperscript{136}

Another instrument, SOPLAY, was developed concurrently with CAST.\textsuperscript{124} This instrument from the US had its criterion validity tested against a PA self report questionnaire administered to middle and high school students.\textsuperscript{124} While self report has acceptable reliability and validity among adolescents, primary school children’s self report of physical activity has not proven sufficiently accurate for validating an instrument.\textsuperscript{127} This problem was avoided in the CAST study where the two versions were tested against expert-scored videos of concurrent field observations and when two concurrent field measurement processes were compared. This strengthened the
case for the validity of direct observation instruments to measure PA engagement levels of large groups in unstructured settings.

One unique aspect of the present study is the fact that reliability and validity measures were determined for different versions of the CAST instrument. The fact that the measures were consistent across these versions adds credibility to the overall assessment system.

When even more CAST data are available it may be possible to apply newly developed time series modelling and signal extraction methods, to ‘borrow information over time’ for improving the estimates. However, these analyses are normally performed on repeated survey data such as labour market participation and annual population health surveys, and there may be sampling issues when applying these methods to CAST data. In her paper to the OECD statistics group, Graf (2003) states that these…“promising but complex methods are not yet applied routinely by statistical agencies”. However, even without conducting such analyses it is clear that findings that show a unidirectional trend over time are more robust than a significant result from a measurement, which was taken only once.

**Relationship between equipment variables and physical activity engagement**
The CAST studies analysed as part of this thesis differed in the way equipment variables were measured and in their relationship to PA engagement levels. While these measures are not central to questions that define the scope of research, it is relevant to briefly outline and discuss the most pertinent findings.

The author had access to field data regarding relationship between equipment variables and PA engagement in three out of the four studies (only reliability data from the Dubbo study were available). In MIGI, the only equipment variable which was a significant predictor of PA was the number of balls in the playground (predicting vigorous PA). No variable, which simply counted equipment items, was a significant predictor of PA engagement in either the Illawarra or Central Coast studies. Unlike MIGI, which used the number of children who were using equipment as a measured variable, these other studies used the percentage of children who were engaged in playing with different equipment items.
As seen in tables 12 and 13 in the results section, the percentage of children who used any sort of equipment other than balls (Frisbees, bats, elastic, ‘rockets’) was a significant predictor of vigorous PA engagement in the Illawarra, and in the Central Coast study the percentage of children playing ball games and of children playing with fixed equipment were significant predictors of PA engagement (moderate and vigorous PA were not distinguished in the Central Coast study). The Central Coast findings are not surprising as using the percentage of children using equipment is a direct utilisation and engagement measure, which is more likely to be associated with increased PA.

One of MIGI’s findings was that school size was a significant predictor of PA engagement, with children in smaller schools being significantly more active (see Table 11, PA engagement decreased linearly on the logit scale with increasing number of students in a school).\textsuperscript{1,2} Similarly, in the Central Coast study, there were less children in the care centres’ playgrounds and children’s usage of equipment had a stronger association with PA engagement than in the MIGI and Illawarra studies which had a mixture of small and large school playgrounds.

The author and his colleagues hypothesised in their paper (2001) that this finding may be due to less choice of play partners leading to reduction of social exclusion of less skilled children and to a greater percentage of children who are included in games.\textsuperscript{1} The Central Coast results, add weight to this hypothesis and it is likely that in large playgrounds fewer children were included in games. Children who were included in a game were playing in larger teams and could be less active while playing. In comparison, in smaller playgrounds, a higher percentage of children were included in games and were more active while playing, as the teams were smaller and there were fewer opportunities to be inactive, even for the less skilled children. Another explanation may be that the increased access to equipment enabled greater choice of activity types, which led to increased engagement in physical activity.
A summary of the different versions of CAST and their features

This thesis has outlined the methodological characteristics of CAST. While some studies describe instruments that have been used according to a research plan, which is time bound and set in advance, this thesis described the evolution of a measurement instrument over a period of nine years. This ‘organic’ development of research has been partly due to the relative newness of the field of children’s physical activity research. This is not uncommon in the field of health promotion evaluation and research, where research instruments sometimes undergo an iterative process of development through their application in a number of different contexts and projects. It is also reflected in health promotion research literature.98, 101, 143

The repeated use of different versions of CAST in the four projects outlined in the thesis provided the author opportunities for in depth analyses of methodological aspects of CAST. However, with the exception of MIGI, all other projects were operated within limited budgets, which limited the scope of the research that was conducted. Despite these limitations, the availability of data from four projects added strength to the instrument’s reliability and validity. It performed well in different contexts, i.e. in diverse geographical locations, as well as in two settings: 1. primary schools, and 2. after-school-hours day care centres.

The table in the following pages summarise the different estimates and measures used. It also outlines other aspects of the different CAST versions including their applicability, advantages, limitations, resources required, the number of testers used and the number of PA categories recorded.
**Table 14: Summary of different versions of CAST, their features, limitations and types of reliability and validity measures used**

<table>
<thead>
<tr>
<th>Project</th>
<th>n categories</th>
<th>n testers</th>
<th>MVPA</th>
<th>VPA</th>
<th>Resources</th>
<th>Applicability</th>
<th>Types of validity tested</th>
<th>Types of reliability tested</th>
<th>Summary of findings, issues and limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td>MIGI</td>
<td>5</td>
<td>5</td>
<td>✓</td>
<td>✓</td>
<td>Resource intensive for training as well as field measurements.</td>
<td>Due to number of testers, better applied in medium to large school playgrounds. Used in 18 schools.</td>
<td>✓ ‘Translational validity’.&lt;sup&gt;158&lt;/sup&gt; Face and content validity in conceptual and formative research stages. ✓ Criterion validity against video of same period.</td>
<td>✓ Consensus estimates  - agreement rate  - Cohen’s Kappa ✓ Consistency estimates  - Cronbach’s alpha  - Pearson Correlation  - 95% CI around means  - ICC field data</td>
<td>High reliability of all consensus and consistency estimates. Large correlation effect for % of children engaged in MVPA and number of children in playground. Only medium correlation strength in validity study of number of children engaged in MVPA. Similar findings when 95% CI around mean were used. The project funding enabled a sophisticated analysis of field, validation and reliability data. This laid the groundwork for further data collection and analytical work in later projects with little or no funding. A number of limitations were rectified in later projects, eg no independent count of overall number of boys/girls in playground. This CAST version provided the most detailed data.</td>
</tr>
<tr>
<td>Project</td>
<td>n categories</td>
<td>n testers</td>
<td>MVPA</td>
<td>VPA</td>
<td>Resources</td>
<td>Applicability</td>
<td>Types of validity tested</td>
<td>Types of reliability tested</td>
<td>Summary of findings, issues and limitations</td>
</tr>
<tr>
<td>---------</td>
<td>-------------</td>
<td>----------</td>
<td>------</td>
<td>-----</td>
<td>-----------</td>
<td>---------------</td>
<td>-----------------------</td>
<td>-----------------------------</td>
<td>-----------------------------------------------</td>
</tr>
<tr>
<td>Illawarra</td>
<td>3</td>
<td>3-4</td>
<td>✓</td>
<td>✓</td>
<td>Less resource intensive than MIGI, but still required 4 raters when independent count of children was done.</td>
<td>Less intrusive in smaller playgrounds. Used in four schools.</td>
<td>✓ ‘Translational validity’ regarding three activity categories. Face and content validity in conceptual and formative research stages. ✓ Criterion validity against video of same period.</td>
<td>✓ Consistency estimates - Cronbach’s alpha - Pearson Correlation - 95% CI around means - ICC field</td>
<td>High reliability of all consistency estimates. High criterion validity. Tested criterion validity against an extra criterion, i.e. number of children in the playground counted by independent fourth observer. More efficient than the CAST version used in MIGI, i.e. provided data regarding both moderate and vigorous physical activity engagement with three instead of five observers. Small number of children in validation study. Small number of scans in both video validation and reliability testing due to limited project funding and size.</td>
</tr>
<tr>
<td>Project</td>
<td>n categories</td>
<td>n testers</td>
<td>MVPA</td>
<td>VPA</td>
<td>Resources</td>
<td>Applicability</td>
<td>Types of validity tested</td>
<td>Types of reliability tested</td>
<td>Summary of findings, issues and limitations</td>
</tr>
<tr>
<td>-----------------</td>
<td>--------------</td>
<td>-----------</td>
<td>------</td>
<td>-----</td>
<td>-----------</td>
<td>--------------------------------------------------------------------------------</td>
<td>------------------------------------------------------------------------------------------</td>
<td>---------------------------------------------------------------------------------------------</td>
<td>----------------------------------------------------------------------------------------------------------</td>
</tr>
</tbody>
</table>
| Central Coast   | 2            | 2-3       | ✓    |     | Less than Illawarra and MIGI. May facilitate a number of observer teams.       | Less intrusive for smaller playgrounds. Used in 6 playgrounds of out of school hours child care centres. | ✓ ‘Translational validity’ regarding two activity categories. Face and content validity in conceptual and formative research stages. ✓ Criterion validity via correlation estimation with independent count of children in playground. No resources in project to test criterion validity against a video. | ✓ Consistency estimates  - Cronbach’s alpha  - Pearson Correlation  - 95% CI around means  - ICC field ✓ Measurement estimates. Factor Analysis (PCA) | High reliability of all consistency estimates.  
Added another setting in which CAST was used reliably.  
Less resource intensive, but with a consequence that there was no distinction between moderate and vigorous PA.  
No validation study comparing to a video, but obtained a high measure of criterion validity comparing with independent count of children. One out of the seven raters was found to be an outlier. |
| Dubbo          | 2            | 2-3       | ✓    |     | Less than Illawarra and MIGI. May facilitate a number of observer teams.       | Used in one large school to evaluate an intensive intervention. | ✓ Criterion validity via correlation estimation with independent count of children in playground. No resources in project to test criterion validity against a video. | ✓ Consistency estimates  - Cronbach’s alpha  - Pearson Correlation  - 95% CI around means | High reliability of all consistency estimates.  
Less resource intensive, but with a consequence that there was no distinction between moderate and vigorous PA.  
No validation study comparing to a video, but obtained a very high measure |
<table>
<thead>
<tr>
<th>Project</th>
<th>n categories</th>
<th>Resources</th>
<th>Applicability</th>
<th>Types of validity tested</th>
<th>Types of reliability tested</th>
<th>Summary of findings, issues and limitations</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>MVPA</td>
<td>VPA</td>
<td></td>
<td></td>
<td>of criterion validity comparing with independent count of children.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Relatively small number of scans for the inter-rater reliability testing.</td>
</tr>
</tbody>
</table>
Limitations

A number of limitations related to the studies outlined in this thesis arose from the applied nature of the research. In a pure or academic research environment it would have been preferable to conduct larger validation studies and apply them to the CAST version, which measured two activity categories in the Central Coast and Dubbo studies. The validation study in the Illawarra project was conducted with a relatively small number of children.

It would have also been better to conduct larger reliability tests where more scans were rated simultaneously by a number of raters. These studies would be easier to achieve with funding which is specifically dedicated to developing the instrument. All projects, which were covered in this thesis, had a very strong health promotion implementation directive, and at times the author had to persuade project workers to dedicate some resources and time to reliability measures.

These limitations stemmed from the nature of health promotion funding, which often includes a modest budget for evaluation purposes that does not allow for extensive instrument development. Health promotion units emphasise implementation over evaluation as they are often required to deliver population health outcomes with a limited budget. Evaluation of programs is seen as important, but costly and lengthy development of research instrument is seen as luxury. This is especially the case with rural health promotion units. Even in the MIGI project, which was funded to create new HP evidence, the funds for the resources dedicated to the playground study were limited by the need to resource the study of two other outcomes and implement the three interventions which targeted those outcomes.²

The author used the opportunity, which writing this thesis provided, to deepen the research of CAST and connect studies that otherwise would have been left as separate, small, and unpublished studies.

Another limitation was that different raters were used in the four studies. One unifying factor was that the author provided training for raters in three of the four studies, and one of the Illawarra raters provided the training for the Dubbo project.
The use of different raters may increase the variance between their ratings as indicated in the Central Coast case. However, since CAST was intended for use in diverse settings and by different organisations, the diversity of raters in the four studies included in this thesis reflects how CAST is likely to be used in the field in future studies. As long as raters are well trained, and reliability measures are collected, analysed and fed back to project coordinators, the data are likely to be reliable.

Another methodological limitation was that the validation and reliability studies were conducted only regarding PA engagement levels. CAST raters collected data on equipment availability, students’ usage of equipment, as well as teachers’ presence and behaviour in the playground. However, in all projects, health promotion workers were more interested in using the limited resources available, to research PA engagement more intensively. There were simply insufficient resources to conduct validation and reliability studies on equipment usage and teacher behaviour variables in the playground. The validity of directly observing some environmental variables in individual children had been established by McKenzie et al..76, 136, 139, 193

Textbook definitions of validity types did not always neatly fit with the exploratory research described in this thesis. The consistency between the above mentioned findings added to the criterion validity of CAST i.e. the degree of correlation with another measure of the same phenomenon.147 It is hard however to clearly define whether it added to concurrent or predictive criterion validity, and in a way it did not neatly fit either of the definitions. Concurrent validity measures the degree of correlation of two measures of the same phenomenon administered at the same time, and predictive validity measures the extent of correlation of a present measure with a future measure.143, 158

The measurements pointed to a similar pattern from different analyses taken at different times, so could not be defined as concurrent. It seems that predictive criterion validity better describes these findings, but unlike the above definition of predictive criterion validity, the measures could not be directly correlated and were compared conceptually after separate regression modelling were performed. Again,
the data and findings in this thesis tended to be strongly influenced by the applied research context, in which they were collected.

One of the main features and limitations of all versions of CAST is that the instrument reliability and validity were tested only for a break level estimate of PA engagement.

PA engagement level of a whole break period is the focus of interest for most researchers and implementers of school playground PA promotion projects, so CAST is a valid and reliable instrument to achieve this end. While one scan is not a valid or reliable measure, the generally high consistency measures, like Cronbach alpha and Pearson correlation coefficients, found in both reliability and validity studies across the four studies, indicate that using CAST can also track the ebb and flow of PA engagement in the playground during the break.

The Pearson correlation coefficient for the number of children engaged in PA, and total number of children in the playground during MIGI and Illawarra validation studies were in the medium-to- large and large correlation size according to Cohen’s interpretation of correlation sizes in psychological research. As Cohen himself has observed, however, all such criteria are in some ways arbitrary and should not be observed too strictly. This is because the interpretation of a correlation coefficient depends on the context and purposes. A correlation of 0.9 may be very low if one is verifying a physical law using high-quality instruments, but may be regarded as very high in the social sciences where there may be a greater contribution from complicating factors. The MIGI video-field rates were around 0.5 while the Illawarra video-field and the three latter studies field-field rates were much higher (0.85-0.98) which may point that it is easier to track the flow of PA engagement using the three and two categories CAST than the five categories CAST due to decreased potential for error when a rater does not need to distinguish between sub-categories within the sedentary category. It might also have been due to the smaller number of children in the playground during the Illawarra video validation study. Another likely explanation is that in MIGI the video and field scans were not synchronised and every four five-second scans in the video were averaged before being compared to the 20-second field scan.
It is important to note that the level of both reliability and validity in all studies covered in this thesis increased as the number of scans during a break increased.\textsuperscript{1,2} CAST would not necessarily be valid or reliable if it was used to assess the PA engagement level of short periods of time or if it was used to assess longer break periods with a small number of scans.

The reliability of a single scan as a measure of a break period varied between the projects and was much higher in the MIGI and Illawarra projects where larger numbers of children were scanned. This reliability is the scan level intra class correlation (ICC) in a multi-level model, with a structure of children within scans within break periods. The ICC of a single scan ranged from 0.13 to 0.34 in the Illawarra and MIGI projects and was much lower in the Central Coast study at 0.08 and 0.04 in the significant variables and null models respectively.

These findings reflected the extent of clustering of, and variance between, different scans as they captured the changes in PA engagement during playtime. The Central Coast data were collected in after-school-hours childcare centres, where measurement periods were longer than recess or lunch breaks in primary schools. A typical recess or lunch period in a primary school consisted of a slow formative few minutes in the beginning of the break (measurement did not start until after the designated eating period) followed by a fairly active period, which came to an abrupt end with the bell at the end of the break. There was less variation in the scans taken in primary school playgrounds than those taken in the childcare centres, which may have waxed and waned according to children’s choice of PA engagement in a less concentrated and energetic way than school break periods.

When the lowest ICC for a scan in the Illawarra study (0.13; model with significant predictors of moderate PA engagement) is used for an average length break (17 scans) the reliability of predicting a break level mean is fairly high at 0.73. The highest scan level ICC of 0.34 was recorded in MIGI for the null model predicting moderate PA engagement. If the average break length of 17 scans were used with this level of ICC, the reliability of predicting the mean PA engagement of a break would be high at 0.9.
While the Central Coast study yielded lower ICCs, the median number of scans per observation period was 40, which provided composite reliabilities of 0.63 and 0.77 to predict break level means of PA engagement in the null and significant variables models respectively.

As seen from the paragraphs above, the ICC at the scan level determines the number of scans required per period to produce a reliable estimate of period level PA engagement. This is explained by Snijders and Boskers (1999) who illustrated that when a level one ICC is 0.1 in a two-level model, it would require 50 level one units to achieve a 0.85 reliability of measuring a level two unit mean. On the other hand, if the level one unit ICC was 0.4, it would only require 10 level one units to achieve a 0.87 reliability of measuring a level two unit mean. In the primary schools studies, 10 scans took just over 12 minutes. Most recess breaks are 15 minutes, while the duration of the non-eating part of lunch is normally 30-40 minutes, which allows for 25-35 scans. Observers may be able to scan more frequently if they are using the simpler CAST version, and especially if they only measure PA engagement and do not scan a second time for equipment variables.

Future users of CAST would be advised to be aware of this limitation and use CAST only when PA engagement levels of whole break periods are sought. It would also be useful to know how clustered or varied scans are, so that if their ICC is high, less scans would be required to reliably measure a break level PA engagement.

The interpretation of confidence intervals around break level means for reliability and validity measures was sometimes difficult. Confidence interval measures should be therefore used to add to and corroborate other measures of reliability and validity. While confidence intervals around period level means of PA engagement may indicate whether there is a difference between two raters’ measurements (or between field and video measures), the number of scans, and therefore the width of confidence intervals around the mean, has an effect on the result. In the MIGI video validation study, there were 186 field scans. This created a very narrow confidence interval around the mean number of children engaged in MVPA, leading to a significant difference in the means for this period as measured by video and field
observers. However, even with this number of scans the means of the number of children in the playground and also of the percentage of children engaged in MVPA were not significantly different between video and field measurements when 95% CI were calculated.

Since the aim of CAST is to accurately estimate the percentage of children who are engaged in PA, the findings indicated that CAST was a valid measure. It was interesting that the Illawarra validity study found no difference between the means of the video and field measures, but used only 50 scans. It has already been noted above that in most breaks only 10-30 scans were taken. Most inter-rater reliability tests in this thesis had numbers of scans within this range so the lack of significant difference between the means may have indicated high reliability rates or may have been due to smaller sample sizes and wider CI’s.

Given this limitation of using means and their CI’s, it was important to include other consistency measures. The usage of both 95% CI around means and other consistency measures (Cronbach’s alpha, ICC’s, Pearson correlation, Factor analysis) provide triangulation of different sources of findings and strengthens the conclusion that CAST is a valid and reliable instrument if used to measure a break level PA engagement.

**Conclusions**

Addressing the main research questions of this thesis, physical activity engagement levels of large numbers of children could be measured in school playgrounds and similar environments using CAST as a direct observation instrument.

CAST validly and reliably measured children’s physical activity levels in school playgrounds. Reliability and validity relied on thorough training of raters as well as follow-up analysis of inter-rater reliability data. CAST validity has been confirmed for the five, three and two categories instruments. Inter-rater reliabilities for all versions (two, three and five categories) were high.
CAST produces its highest quality data when the two or three categories versions are used to observe large groups in well-defined playground. A number of scans are required to produce a reliable estimate of a break level physical activity engagement so scanning regularly throughout a typical school break period is preferable to sampling a small number of scans. This number depends on how clustered the scan level values are, that is the similarity between scans.

**Recommendations for further research**

Considering the ‘accuracy-practicality trade-off’ mentioned in this thesis and by Miller and Welk\textsuperscript{127, 179} It would be useful if, in future implementation research on playground PA engagement, the measurement methods had low costs in terms of human and financial resources required.

Building on the work outlined in this thesis, more research is needed to further validate the two main components of CAST: 1. engagement levels in different PA levels (with less than three observers), and 2. equipment availability, equipment usage and teacher presence.

As resources are often limited in both health promotion and educational research studies, it would be useful to assess the instrument’s ability to accurately measure how many children are sedentary and how many are active (without differentiating whether they are moderately or vigorously active) when two observers are scanning. If resources for such a study were available, this could be done by comparison with video analysis, accelerometers or another version of CAST.

In order to enable even more economic use of CAST, It would be useful to develop and test a CAST version for one observer. Piloting by training observers to assess PA engagement using a one observer instrument should include a validation study against either a video or data obtained by a larger CAST team (of two or three observers) scanning the same area at the same time. A one person CAST will entail the observer scanning the playground for the number of sedentary boys or girls, followed by a scan to count the overall number of boys or girls in the playground. A one person CAST may need to be tested in different size playgrounds, as it may be
applicable only in small to medium playgrounds. It is unlikely that a one person CAST would enable data collection regarding equipment usage and teacher presence.

Another validation study may be conducted regarding the validity of data collection on equipment availability, equipment usage and teacher presence. It would be useful to conduct separate validation studies for four three and two observers CAST versions, and determine what version of CAST yields valid measurement of these variables.

It would also be useful to conduct a study regarding the ability of school children to use CAST reliably. Such a study can determine the ability of children of different ages to use CAST validly and reliably.


63. MAYO Clinic. Obesity Researchers Test “Classroom of the Future”. In: Newswise; 2006.
<table>
<thead>
<tr>
<th>No.</th>
<th>Reference</th>
</tr>
</thead>
</table>


List of Appendices

Appendix 1: CAST training manual
Appendix 2: Preventive Medicine paper regarding the MIGI Project
Appendix 3: Health Promotion Journal of Australia brief report - Illawarra
Appendix 4: PA engagement results of the Illawarra project playground intervention
Appendix 5: PA engagement results of the Central Coast project
Appendix 6: MIGI project evaluation summary report
Appendix 7: A paper about the use of CAST in another project
CAST

Children Activity Scanning Tool

A training manual for the Central Coast AHS Physical Activity in Out Of School Hours (OOSH) care centres project

Prepared by Avigdor Zask
Research & Evaluation Officer
Health Promotion Unit
Northern Rivers Area Health Service
Lismore
NSW 2480
Australia

avigdor@nrhs.health.nsw.gov.au

August 2003
1. Categories of physical activity (PA) used in CAST – the SOFIT instrument and its modification. (This section is based on McKenzie 1998).

SOFIT (System for Observing Fitness Instruction Time)
SOFIT is an objective tool for assessing the quality of physical education instruction (1). It is a comprehensive system that provides a measure of student activity, lesson context, and teacher behavior during class time.

SOFIT involves the direct observation of lessons by trained observers and has been used to assess physical education in over 300 schools throughout the United States. These include the CATCH (3, 4, 5) M-SPAN, and SPARK (2, 8) Projects which are three intervention studies supported by the National Institutes of Health. The main focus of SOFIT is on the coding of student physical activity levels and selected environmental factors (lesson context and teacher behavior) that are associated with opportunities for students to be physically active and become physically fit.

SOFIT student physical activity codes have been validated by both heart rate monitors (1, 10) and CALTRAC accelerometers (6). Lesson context categories have been developed from definitions used commonly in physical education evaluation research (9). Reliabilities for independent trained observers have consistently exceeded 90% agreement on all SOFIT categories, which indicates the measures are accurate (1, 2, 4, 5, 8).

SOFIT enables teachers and researchers to make judgments about physical education lessons, particularly as they relate to program goals.

The SOFIT System - Technical Description

Student physical activity engagement.
A decision regarding the physical activity levels of individual students is made. Each student’s involvement decision determines his/her level of physical activity (active engagement level). The engagement level provides an estimate of the intensity of the student's physical activity. While the CATCH (3, 5) and MIGI (11) programs used 5 activity codes, it is proposed that the CCAHS playground activity project will amalgamate codes 1, 2 and 3 into a ‘non-active’ level while retaining the codes for moderate PA (walking) and running (vigorous PA). The higher the code, the higher the student's rate of energy expenditure. See table below for a quick reference to activities and their corresponding codes.

In all the projects mentioned above SOFIT was used to gauge PA levels in PE classes and other aspects of PE classes like lesson context were also recorded. However, for the CCAHS project SOFIT categories will only form the basis to using the CAST instrument in OOSH care centres. Thus, a decision regarding the activity level of each student in the centre will be made by every observer as they scan the playground or rooms. See the following table for a summary of PA levels used in CAST:
### CAST CATEGORY

<table>
<thead>
<tr>
<th>CAST CATEGORY &amp; EXCEPTIONS</th>
<th>1 (Lying)</th>
<th>1 (Sitting)</th>
<th>1 (Standing)</th>
<th>2 (Walking)</th>
<th>3 (Running)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rolling 2/3</td>
<td>Sliding 1/2/3</td>
<td>Into squat 2</td>
<td>Skip 3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tumbling 2/3</td>
<td>Bouncing 2/3</td>
<td>Up from squat 2</td>
<td>Shuffling 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Push ups 3</td>
<td>Into squat 2</td>
<td>On tip toes 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sit ups 2/3</td>
<td>Throw ball</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
2. Using CAST in the field

CAST is an instrument that uses the SOFIT categories to assess PA levels of children in school playgrounds. It was developed in the NRAHS as part of the MIGI project (11). This manual will explain a new version of CAST to be used by Central Coast Area Health Service personnel in their 2003 PA in Out of School (OOSH) care centers pilot project.

The following is an extract from a paper published on the use of CAST (13, pp 403-404):

“The Children activity scanning tool (CAST) was developed to assess PA levels, equipment availability/use and teacher presence/behaviour during school break times of recess and lunch. It used a team of five observers to simultaneously and repeatedly scan a play area every 75 seconds (by audio taped signal) until the break ended. For full viewing coverage, each playground was divided into discrete areas which were given equal scanning time. For each scan, all observers simultaneously swept the area visually in the same direction. All schools had a designated eating time at the start of the lunch break that was not included in the observation.

Scans alternately focused on boys and girls. The task of each observer was to first scan the designated area for the number of boys or girls engaged in one of five PA levels according to the SOFIT instrument…. Each observer was also allocated an equipment or teacher category to observe and record on their second sweep. Equipment categories were: number of balls in the area, number of children playing ball games i.e. focused on a ball (eye contact and body language/direction) and/or actively manipulating the ball, number of children playing with non-fixed equipment other than balls, number of children playing on fixed equipment. Teacher behaviour/presence categories were: number of teachers present in area, number of teachers encouraging PA i.e. verbal encouragement/feedback or teacher’s participation in activities, number of teachers observing including passive supervision, number of teachers managing including discipline, allocating equipment and active involvement in playground supervision.

Prior to each scanning session, temperature and humidity (by hygrometer; for heat stress level) and ‘wet’ or ‘dry’ day (‘wet’ if evidence of rain prior to or during observation) were recorded along with numbers of available equipment items…”

CAST will be used in a similar way in CCAHS, but only 3 PA categories will be recorded. The team’s size can be 3 or 4 observers. If 4 observers are used, 3 observers can record the 3 new categories of PA while 1 observer can count the total number of children in the playground. After finishing the first sweep measuring the PA levels, a second sweep for equipment usage and availability and teacher behaviour can be undertaken. Intervals between scans may need to be longer because the observation team will be smaller, but will still need to record equipment usage and teacher behaviour (which may require another sweep in each scan).
procedure for scanning equipment usage/availability and teacher behaviour will be decided at the training with CCAHS personnel depending on their project’s goals.

**Using the scoring sheet**

Good quality data is essential to the evaluation of your project. As part of the team you can ensure that quality is achieved and maintained. Please don’t leave any fields blank on the front page of the scoring sheet unless instructed to do so by your team leader. (Eg: If you check and can see no monkey bars then enter a zero to indicate that you did check. A blank will make it difficult to know what actually occurred). Please don’t leave the site until your team leader has checked that your records are complete and readable.

If you accidentally enter a wrong number, make sure that the correct entry is readable. (It may be clearer to put the correct value in a margin with an arrow pointing to the original spaces rather than try to squash it in where there is not enough space).

Let’s go through fields which may need clarification.

- **Total kids attending on observation day**: Get this information from the OOSH coordinator on arrival.

- **Total carers on duty**: Obtain this information from the OOSH coordinator on arrival. Include all permanent and casual carers and volunteers.

- **Starting and finishing times**: Each team is to arrive at the OOSH centre at 3.45 pm and commence observing at 4pm. An indoor observation and an outdoor observation is to be done at each centre. You may want to comment if scanning stopped eg we ran out of batteries/interrupted by rain. Check that all members of your team have recorded the same time.

- **Observer Team Identifier**: String together all first and last name initials of all team members.

- **Wet or dry**: If there is evidence of there having been rain earlier on the day or if it is raining at the time of observation then circle ‘wet’ otherwise circle ‘dry’.  

- **Humidity**: These are taken from the wet/dry thermometer apparatus as described in the attached ‘Wet Globe Thermometer Procedures’

- **N children by CAST category**: Record number of boys or girls observed in each scan in the activity level which has been assigned to you.

- **No. balls**: Total number of balls in the playground on this scan.

- **No. play ball**: Number of kids of gender being scanned (boy/girl) who are engaged in a ball-based activity.
**No. play equipment** Number of kids of gender being scanned (boys/girl) who are engaged in an activity using non-fixed equipment.

**No. play equipment** Number of kids of gender being scanned (boys/girl) who are engaged in an activity using fixed equipment.

**Teacher behaviour** Enter the number of teachers engaged in each behaviour on this scan.

See a sample scoring sheet below. As well as PA levels, equipment usage, equipment availability and teachers’ behaviours are also recorded. Numbers will be entered into all boxes eg number of boys/girls in a certain PA category, number of boys/girls playing with balls, number of teachers who are observing. Please note that unlike all other entries which alternate between boys and girls, the number of balls in the whole playground will be recorded each time.

---

**Children Activity Scanning Tool (CAST) Scoring Sheet**

<table>
<thead>
<tr>
<th>Date:</th>
<th>Centre name:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total kids attending:</td>
<td>Total number of carers on duty:</td>
</tr>
<tr>
<td>Time scanning started:</td>
<td>Time scanning ended:</td>
</tr>
<tr>
<td>Observer team identifier:</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Wet □ Dry □</th>
<th>Temp dry bulb:</th>
<th>Temp wet bulb:</th>
<th>Humidity:</th>
</tr>
</thead>
</table>

**N fixed equipment components:**

<table>
<thead>
<tr>
<th>Painted ‘targets’ on wall:</th>
<th>Painted ‘targets’ on the ground:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monkey bars:</td>
<td>Slippery dips:</td>
</tr>
<tr>
<td>Netball hoops:</td>
<td>Basket ball boards/hoops:</td>
</tr>
<tr>
<td></td>
<td>Other:</td>
</tr>
<tr>
<td>Scans</td>
<td>N of children in each category</td>
</tr>
<tr>
<td>-------</td>
<td>--------------------------------</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>BOYS1</td>
<td></td>
</tr>
<tr>
<td>GIRLS1</td>
<td></td>
</tr>
<tr>
<td>BOYS2</td>
<td></td>
</tr>
<tr>
<td>GIRLS2</td>
<td></td>
</tr>
<tr>
<td>BOYS3</td>
<td></td>
</tr>
<tr>
<td>GIRLS3</td>
<td></td>
</tr>
<tr>
<td>BOYS4</td>
<td></td>
</tr>
<tr>
<td>GIRLS4</td>
<td></td>
</tr>
<tr>
<td>BOYS5</td>
<td></td>
</tr>
<tr>
<td>GIRLS5</td>
<td></td>
</tr>
<tr>
<td>BOYS6</td>
<td></td>
</tr>
<tr>
<td>GIRLS6</td>
<td></td>
</tr>
<tr>
<td>BOYS7</td>
<td></td>
</tr>
<tr>
<td>GIRLS7</td>
<td></td>
</tr>
<tr>
<td>BOYS8</td>
<td></td>
</tr>
<tr>
<td>GIRLS8</td>
<td></td>
</tr>
<tr>
<td>BOYS9</td>
<td></td>
</tr>
<tr>
<td>GIRLS9</td>
<td></td>
</tr>
<tr>
<td>BOYS10</td>
<td></td>
</tr>
<tr>
<td>GIRLS10</td>
<td></td>
</tr>
<tr>
<td>BOYS11</td>
<td></td>
</tr>
<tr>
<td>GIRLS11</td>
<td></td>
</tr>
<tr>
<td>BOYS12</td>
<td></td>
</tr>
<tr>
<td>GIRLS12</td>
<td></td>
</tr>
</tbody>
</table>
3. Observation procedures
The following section describes what has to be done before and after observation. These procedures have been written up as checklists so observers can tick each task as it is completed.

CAST Team Task List

<table>
<thead>
<tr>
<th>Task</th>
<th>Done</th>
<th>Not Done</th>
</tr>
</thead>
<tbody>
<tr>
<td>Make sure the OOSH coordinator knows about the observation by ringing the day before you leave (Friday if your visit is on Monday).</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Get to the centre at 3.30pm</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Notify the OOSH coordinator that you have arrived (you may need to sign the centre’s visitors book)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fill in the centre and observation details</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Check when afternoon tea is (normally between 3 &amp; 3.45) and ensure you do not observe within this period</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Allocate who observes which categories and circle them for each observer (do not use highlighter)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ensure every observer has enough blank forms</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ensure your team has a copy of the centre site map.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Walk around the whole centre area checking for potential hidden play areas</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Decide on observation positions and movement between them. Mark locations on the site map. Make sure you spend an even amount of time in each 'vantage point' and allow time to move between them. A minimum of 2x 20 minutes obs, 1 indoor &amp; 1 outdoor will be taken in each OOSH CC.</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Designate a time keeper/announcer</td>
<td></td>
<td></td>
</tr>
<tr>
<td>The time keeper keeps a walkman on and calls “now” when s/he hears the sound</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Decide direction of ‘sweeping’ the areas you observe. Mark it on the centre’s site map using arrows.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Collect all record forms at the end of the observation and enter into the raw data folder</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Check and record number of children attending centre on observation day.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Some centres have organised things like piano lessons, karate that only some kids attend. It is preferable to observe when such activities are not scheduled.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Things to note

| Once you start counting children in the ‘sweeping’ direction, don’t go back |
| The number of children participating in ball games (No. play ball) is the hardest to determine |
| Children that are waiting their turn to do an activity within a game are counted even if they are not currently hitting/batting etc. |
| Any children in the game area whose body language suggests they are part of the game (i.e. turning towards where ball is, eye contact) are counted even if they are currently passive. If you are at all uncertain whether they are participating in a game, do not count them. |

If you are asked what you are doing by a child or a teacher always give the following answers:

- **Carers:** Tell them you are observing physical activity in the centre. If asked further, say you are looking at what activities children are engaged in. If they want more information (which is very unlikely), please give them the project officers’ contact numbers.
- **Children:** Tell them we are looking at what kids do in the playground or in their play time.
References


Appendix 2: Preventive Medicine paper regarding the MIGI project playground component
Appendix 3: Health Promotion Journal of Australia brief report regarding the Illawarra project
Appendix 4: Physical activity engagement results of the Illawarra project

Opposing trends were found when Moderate and Vigorous PA levels before and after the intervention were compared. For schools 1, 3 and 4, an increase in MPA level was matched by a decrease in VPA. School 2 had no changes in MPA following the intervention and a decrease in boys’ VPA, but an increase in girls’ VPA. These findings indicate that the intervention did not produce a significant increase in the proportion of children who are engaged in moderate or vigorous physical activity in the school playgrounds observed.

Overall, significantly less girls were engaged in both MPA and VPA. Significantly more boys were vigorously active during lunch breaks than during recess in all schools. Table 1 shows the percentage of children engaged in MPA and VPA under different levels of predictor variables.

**Table 1:** Estimated mean percentages MVPA and VPA and 95% confidence intervals for different levels of predictor variables in the Illawarra project.

<table>
<thead>
<tr>
<th>School</th>
<th>Gender</th>
<th>Pre/Post</th>
<th>%Children in MPA</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Boy</td>
<td>Pre</td>
<td>31.1 (29.6-32.6)</td>
</tr>
<tr>
<td>1</td>
<td>Boy</td>
<td>Post</td>
<td>36.4 (34.8-38.0)</td>
</tr>
<tr>
<td>1</td>
<td>Girl</td>
<td>Pre</td>
<td>26.9 (25.6-28.3)</td>
</tr>
<tr>
<td>1</td>
<td>Girl</td>
<td>Post</td>
<td>31.9 (30.4-33.4)</td>
</tr>
<tr>
<td>2</td>
<td>Boy</td>
<td>Pre</td>
<td>43.9 (42.2-45.6)</td>
</tr>
<tr>
<td>2</td>
<td>Boy</td>
<td>Post</td>
<td>43.7 (42.0-45.5)</td>
</tr>
<tr>
<td>2</td>
<td>Girl</td>
<td>Pre</td>
<td>39.0 (37.4-40.7)</td>
</tr>
<tr>
<td>2</td>
<td>Girl</td>
<td>Post</td>
<td>38.9 (37.2-40.6)</td>
</tr>
<tr>
<td>3</td>
<td>Boy</td>
<td>Pre</td>
<td>45.1 (43.4-46.8)</td>
</tr>
<tr>
<td>3</td>
<td>Boy</td>
<td>Post</td>
<td>36.9 (35.2-38.5)</td>
</tr>
<tr>
<td>3</td>
<td>Girl</td>
<td>Pre</td>
<td>40.2 (38.5-41.9)</td>
</tr>
<tr>
<td>3</td>
<td>Girl</td>
<td>Post</td>
<td>32.3 (30.8-33.9)</td>
</tr>
<tr>
<td>4</td>
<td>Boy</td>
<td>Pre</td>
<td>31.1 (29.6-32.6)</td>
</tr>
<tr>
<td>4</td>
<td>Boy</td>
<td>Post</td>
<td>40.4 (38.7-42.1)</td>
</tr>
<tr>
<td>4</td>
<td>Girl</td>
<td>Pre</td>
<td>26.9 (25.6-28.3)</td>
</tr>
<tr>
<td>4</td>
<td>Girl</td>
<td>Post</td>
<td>35.7 (34.1-37.3)</td>
</tr>
<tr>
<td>School Gender</td>
<td>Pre/Post</td>
<td>%VPA Recess No children playing with equipment</td>
<td>%VPA Lunch No children playing with equipment</td>
</tr>
<tr>
<td>---------------</td>
<td>---------</td>
<td>------------------------------------------------</td>
<td>---------------------------------------------</td>
</tr>
<tr>
<td>1 Boy Pre</td>
<td>16.0 (13.3-19.2)</td>
<td>18.9 (15.7-22.5)</td>
<td>20.1 (16.8-23.9)</td>
</tr>
<tr>
<td>1 Boy Post</td>
<td>9.1 (7.4-11.1)</td>
<td>11.2 (9.2-13.6)</td>
<td>13.4 (11.0-16.2)</td>
</tr>
<tr>
<td>1 Girl Pre</td>
<td>6.2 (5.1-7.7)</td>
<td>7.5 (6.1-9.2)</td>
<td>8.0 (6.5-9.8)</td>
</tr>
<tr>
<td>1 Girl Post</td>
<td>16.0 (13.3-19.2)</td>
<td>18.9 (15.7-22.5)</td>
<td>20.1 (16.8-23.9)</td>
</tr>
<tr>
<td>2 Boy Pre</td>
<td>9.1 (7.4-11.1)</td>
<td>11.2 (9.2-13.6)</td>
<td>13.0 (10.7-15.7)</td>
</tr>
<tr>
<td>2 Girl Pre</td>
<td>13.0 (10.7-15.7)</td>
<td>18.2 (15.1-21.7)</td>
<td>16.4 (13.5-19.7)</td>
</tr>
<tr>
<td>2 Girl Post</td>
<td>10.7 (8.8-13.0)</td>
<td>11.5 (9.4-13.9)</td>
<td>13.6 (11.2-16.5)</td>
</tr>
<tr>
<td>3 Boy Pre</td>
<td>6.1 (5.0-7.5)</td>
<td>8.9 (7.2-10.8)</td>
<td>9.7 (6.4-9.7)</td>
</tr>
<tr>
<td>3 Boy Post</td>
<td>16.0 (13.3-19.2)</td>
<td>18.9 (15.7-22.5)</td>
<td>20.1 (16.8-23.9)</td>
</tr>
<tr>
<td>3 Girl Pre</td>
<td>11.1 (9.1-13.4)</td>
<td>15.6 (12.9-18.8)</td>
<td>14.1 (11.6-17.0)</td>
</tr>
<tr>
<td>3 Girl Post</td>
<td>11.9 (9.8-14.4)</td>
<td>14.1 (11.7-17.0)</td>
<td>15.1 (12.5-18.2)</td>
</tr>
<tr>
<td>4 Boy Pre</td>
<td>16.0 (13.3-19.2)</td>
<td>18.9 (15.7-22.5)</td>
<td>20.1 (16.8-23.9)</td>
</tr>
<tr>
<td>4 Boy Post</td>
<td>11.2 (9.2-13.6)</td>
<td>15.9 (13.1-19.0)</td>
<td>14.3 (11.8-17.2)</td>
</tr>
<tr>
<td>4 Girl Pre</td>
<td>8.2 (6.7-10.1)</td>
<td>11.8 (9.7-14.3)</td>
<td>10.6 (8.6-12.8)</td>
</tr>
<tr>
<td>4 Girl Post</td>
<td>11.9 (9.8-14.4)</td>
<td>14.1 (11.7-17.0)</td>
<td>15.1 (12.5-18.2)</td>
</tr>
</tbody>
</table>
Models’ results are reported in Table 2. Parameter estimates are considered to be significant (two-tailed at p=.05) if they are 1.96 or more times their standard errors.

**Table 2**: Parameter estimates and standard errors from variance components and significant effects only for MVPA and VPA in the Illawarra project.

<table>
<thead>
<tr>
<th>Fixed effects</th>
<th>MPA Null Model¹</th>
<th>MPA Predictor Model²</th>
<th>VPA Null Model¹</th>
<th>VPA Predictor Model²</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coeff. S.E.</td>
<td>Coeff. S.E.</td>
<td>Coeff. S.E.</td>
<td>Coeff. S.E.</td>
</tr>
<tr>
<td>Intercept</td>
<td>-0.585 0.020</td>
<td>-0.821 0.043</td>
<td>-1.749 0.029</td>
<td>-1.657 0.113</td>
</tr>
<tr>
<td>Prepost³</td>
<td>0.253 0.060</td>
<td>-0.646 0.083</td>
<td></td>
<td>-0.410 0.054</td>
</tr>
<tr>
<td>Gender⁴</td>
<td>-0.189 0.038</td>
<td>-0.410 0.054</td>
<td></td>
<td>-0.410 0.054</td>
</tr>
<tr>
<td>School2Dum⁵</td>
<td>0.531 0.069</td>
<td>-0.663 0.102</td>
<td></td>
<td>-0.663 0.102</td>
</tr>
<tr>
<td>School3Dum⁵</td>
<td>0.600 0.068</td>
<td>-0.663 0.102</td>
<td></td>
<td>-0.663 0.102</td>
</tr>
<tr>
<td>School4Dum⁵</td>
<td></td>
<td>-0.663 0.102</td>
<td></td>
<td>-0.663 0.102</td>
</tr>
<tr>
<td>Prepost*Sch2⁶</td>
<td>-0.257 0.098</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prepost*Sch3⁶</td>
<td>-0.609 0.100</td>
<td>1.291 0.142</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prepost*Sch4⁶</td>
<td>0.169 0.070</td>
<td>0.300 0.095</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Girl<em>PP</em>Sch2⁷</td>
<td>0.810 0.168</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Girl<em>PP</em>Sch3⁷</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Girl<em>PP</em>Sch4⁷</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Break/Recess⁸</td>
<td>0.199 0.059</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ProplayBall⁹</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ProplayEquip⁹</td>
<td>0.550 0.220</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adj Heat stress¹⁰</td>
<td>0.166 0.024</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Random effects¹¹</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scan</td>
<td>0.221 0.018</td>
<td>0.155 0.015</td>
<td>0.459 0.035</td>
<td>0.305 0.028</td>
</tr>
<tr>
<td>Child¹²</td>
<td>1.000 0.000</td>
<td>1.000 0.000</td>
<td>1.000 0.000</td>
<td>1.000 0.000</td>
</tr>
<tr>
<td>Statistics:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ICC=r¹³</td>
<td>0.181</td>
<td>0.315</td>
<td></td>
<td></td>
</tr>
<tr>
<td>r²¹⁴</td>
<td>0.790</td>
<td>0.886</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R²¹⁵</td>
<td>0.054</td>
<td>0.106</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

¹ Variance components model
Model with significant and non significant variables. Only significant variables and interactions estimates are included.

Pre or post intervention measurement (reference category = pre)

Dummy code to represent child gender (reference category = male)

Dummy code for schools 2-4 (reference category = school number 1, schools 1&2 intervention, Schools 3&4 control)

Dummy code for interaction variables between Pre/Post and school number (reference category = school number 1 at pre)

Dummy code for interaction variables between gender, Pre/Post and school number 4 (reference category = boys Pre intervention in school number 1)

Variable indicating break type (recess or lunch) (reference category = recess) heat stress

Variables measuring the number of children playing with balls or other equipment (eg ropes, frisbees, bats, but excluding fixed equipment) as a proportion of total number of children scanned

Adjusted Heat Stress = Heat Stress level deviation from the mean (Mean HSL = 76.6). Using adjusted Heat Stress level instead of Heat Stress level makes it much easier to understand the models’ results.

Variance components and their standard errors (SE)

These values are default, given that observations on children within scans provide scan logits

ICC= Intra Class Correlation = \( r_1 \) = reliability of a single scan as a measure of a school

\( r_1 \) = reliability of mean of 17 scans (the median number of scans per break) as a measure of a school

R\(^2\) = proportion of scan variance accounted for by the model
Appendix 5: Physical activity engagement results of the Central Coast project

There were no significant intervention effects on MVPA levels. After this finding, a model with all other explanatory variables was fitted. The only variables that were significant were equipment usage and child-care worker presence. This was the only study in which gender was not a significant predictor of MVPA engagement. The proportion of children who were playing with balls, and the proportion of children who were playing with fixed equipment, were positively correlated with MVPA engagement. The number of workers present was negatively correlated to the percentage of children who were engaged in MVPA. Table 1 shows the percentage of children engaged in MVPA under different levels of predictor variables.

Table 1: Estimated mean percentages of MVPA engagement for different scenarios of predictor variables prevalence (lower and upper 95% confidence intervals).

<table>
<thead>
<tr>
<th>N workers present in playground</th>
<th>% of children engaged in ball games</th>
<th>% of children playing with fixed equipment</th>
<th>% of children engaged in MVPA</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>37.5 (32.6-42.5)</td>
</tr>
<tr>
<td>0</td>
<td>25</td>
<td>0</td>
<td>46.6 (41.4-51.8)</td>
</tr>
<tr>
<td>0</td>
<td>25</td>
<td>25</td>
<td>49.8 (44.5-55.0)</td>
</tr>
<tr>
<td>0</td>
<td>50</td>
<td>0</td>
<td>55.9 (50.6-61.0)</td>
</tr>
<tr>
<td>0</td>
<td>50</td>
<td>25</td>
<td>59.0 (53.8-64.0)</td>
</tr>
<tr>
<td>0</td>
<td>50</td>
<td>50</td>
<td>62.1 (57.0-66.9)</td>
</tr>
<tr>
<td>0</td>
<td>75</td>
<td>0</td>
<td>64.8 (59.9-69.5)</td>
</tr>
<tr>
<td>0</td>
<td>75</td>
<td>25</td>
<td>67.7 (62.9-72.2)</td>
</tr>
<tr>
<td>0</td>
<td>100</td>
<td>0</td>
<td>72.9 (68.5-76.8)</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>34.8 (30.1-39.7)</td>
</tr>
<tr>
<td>1</td>
<td>25</td>
<td>0</td>
<td>43.7 (38.6-48.9)</td>
</tr>
<tr>
<td>1</td>
<td>25</td>
<td>25</td>
<td>46.9 (41.6-52.1)</td>
</tr>
<tr>
<td>1</td>
<td>50</td>
<td>0</td>
<td>53.0 (47.7-58.2)</td>
</tr>
<tr>
<td>1</td>
<td>50</td>
<td>25</td>
<td>56.2 (50.9-61.3)</td>
</tr>
<tr>
<td>1</td>
<td>50</td>
<td>50</td>
<td>59.3 (54.1-64.3)</td>
</tr>
<tr>
<td>1</td>
<td>75</td>
<td>0</td>
<td>62.2 (57.1-67.0)</td>
</tr>
<tr>
<td>1</td>
<td>75</td>
<td>25</td>
<td>65.1 (60.2-69.8)</td>
</tr>
<tr>
<td>1</td>
<td>100</td>
<td>0</td>
<td>70.5 (65.9-74.7)</td>
</tr>
</tbody>
</table>

Model results are reported in Table 2. Parameter estimates are considered to be significant (two-tailed at p=.05) if they are 1.96 or more times their standard errors.
Table 2: Parameter estimates and standard errors from variance component and significant effects only models for MVPA

<table>
<thead>
<tr>
<th>Fixed Effects</th>
<th>Model 1 (variance components only)</th>
<th>Model 2 (significant variables only)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coefficient</td>
<td>Std Error</td>
</tr>
<tr>
<td>$\beta_{1jk} = \text{Intercept}$</td>
<td>-0.336</td>
<td>0.063</td>
</tr>
<tr>
<td>$\beta_{2jk} = % \text{ children playing ball games}$</td>
<td>1.500</td>
<td>0.148</td>
</tr>
<tr>
<td>$\beta_{3jk} = % \text{ children playing with fixed equipment}$</td>
<td>0.512</td>
<td>0.127</td>
</tr>
<tr>
<td>$\beta_{4jk} = n \text{ teachers present in playground}$</td>
<td>-0.116</td>
<td>0.049</td>
</tr>
</tbody>
</table>

Random Effects

<table>
<thead>
<tr>
<th></th>
<th>Model 1 (variance components only)</th>
<th>Model 2 (significant variables only)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$v_{1k} = \text{school}$</td>
<td>0.015</td>
<td>0.014</td>
</tr>
<tr>
<td>$u_{2jk} = \text{scan}$</td>
<td>0.353</td>
<td>0.049</td>
</tr>
<tr>
<td>$e_{0jk} = \text{child}^1$</td>
<td>1.000</td>
<td>0.000</td>
</tr>
</tbody>
</table>

Statistics

<table>
<thead>
<tr>
<th></th>
<th>Model 1 (variance components only)</th>
<th>Model 2 (significant variables only)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ICC = $r_1^2$</td>
<td>0.041</td>
<td>0.077</td>
</tr>
<tr>
<td>$r_2^3$</td>
<td>0.630</td>
<td>0.769</td>
</tr>
<tr>
<td>$R_1^2$</td>
<td>0.435</td>
<td></td>
</tr>
<tr>
<td>$R_2^2$</td>
<td>0.117</td>
<td></td>
</tr>
</tbody>
</table>

$^1$ These values are default given that observations on children within scans provide scan level logits

$^2$ ICC = intraclass correlation coefficient = $r_1$ = reliability of a single scan as measure of a school

$^3$ $r_2$ = reliability of mean of 40 (median number of scans per break) scans as measure of a centre

$^4$ $R_1^2$ = proportion of scan variance accounted for

$^5$ $R_2^2$ = proportion of centre variance accounted for
Appendix 6: MIGI project evaluation summary report

Only the cover sheets, and one page of the executive summary, which mentions CAST are included. The full report can be found at:
Appendix 7: A British Journal of School Nursing paper about another project in which CAST was used.

The author could not get permission to reproduce this paper as part of the thesis. If you are interested to read the full paper, please obtain it via the journal: