


# Physical Education lesson intensity and complexity affect subsequent on-task behaviour in the primary school classroom

DPhil Education
January 2020

Christina Hubertina Helena Maria Heemskerk

Candidate number 773720
Green Templeton College

# Physical Education lesson intensity and complexity affect subsequent on-task behaviour in the primary school classroom

Christina Hubertina Helena Maria Heemskerk

# Declaration

I certify that this thesis is a true and honest account of my own research, which was conducted ethically and the results obtained are genuine. Where I have drawn on the work, ideas, and results of others, this has been appropriately acknowledged in this thesis.

Content from this thesis has been accepted for publication in the *Journal of Science and Medicine in Sport*, and is available online. It has been deposited in the Oxford Research Archive (ORA) and will be available open access once the 12-month embargo period has concluded.

> **Heemskerk, C.**, Lubans, D., Strand, S. & Malmberg, L.-E. (in press) *'The effect of physical education lesson intensity and cognitive demand on subsequent learning behaviour'*. Journal of Science and Medicine in Sport. DOI: https://doi.org/10.1016/j.jsams.2019.12.012.

CRediT author statement

**Christina Heemskerk:** Conceptualisation, Data curation, Formal analysis, Investigation, Methodology, Visualisation, Writing – original draft. **David Lubans:** Supervision, Writing – review & editing. **Steve Strand:** Supervision, Writing – review & editing. **Lars-Erik Malmberg:** Supervision, Writing – review & editing.

# Acknowledgements

There is a long list of people who have made this study possible, and whose support over the past three-and-a-half years has been invaluable. There would have been no DPhil for me to write this thesis about if I had not had the support from all the schools, teachers, and participants in my study. Their willingness to accommodate my presence in their lessons, wear the devices, and answer questionnaires has been vital to make this study possible. For that, I would like to thank them first and foremost.

The next thanks go to my supervisors, professors Lars-Erik Malmberg and Steve Strand, who guided my work, challenged me to extend my coding skills, and showed endless confidence in me, especially at the moments when I needed it most. Meanwhile, they also encouraged me to put my own research topic into practice over the course of my studies, and continue my own sports pursuits throughout. Importantly, they supported me in my application for the ESRC's Overseas Institutional Visit funding to work at the University of Newcastle, Australia, for three months. This visit was a defining period in my experience of the DPhil.

I would also like to mention all the staff at the Priority Research Centre for Physical Activity and Nutrition at the University of Newcastle, Australia, and in particular the supervisor for my visit there: Professor David Lubans. Getting involved in their project, the support for my work, the people I was introduced to, and the events and conferences I learned about have all been instrumental in the development of my own study and thinking processes.

A very special thank you Mirna, for many afternoons spent in school with me to collect data. For all the laughs along the way, especially at the most stressful times, and your tireless work, even extending your own term time to match the schools we were working in. It's beyond words how much of a difference this has made to my study, and I can't wait to see you complete your own DPhil!!

Finally, I would like to thank everybody who has listened to my thoughts and ideas, discussed them with me, and challenged my thinking, preferably whilst sharing a meal, coffee, tea, wine, and/or baked goods with me. The list is too long to mention everyone, but special mentions in this area go to Jolet for many meals, coffees, and teas, as well as much shared experience and advice. To Steph, for all the running, breakfast, baking, meals, tea, and laughs. Also to Sina, for much more running; combining physical activity, discussions, and lots of fresh air. To Hanneke and Nick, for much and freely shared wisdom. To Arwen, for incredible proof-reading of large volumes of text at short notice. My family for all the support messages at important moments. And especially to my niece and nephew, Floor and Cas, for giving up their desks so that I could continue working during family visits. Of course, to Marley, for taking me out in the fresh air at least twice a day, joining me on all my runs, and being by my side during long days of coding and writing. And, finally, to everyone who has looked after him to enable me to attend courses, conferences, and data collection days.

# Table of Contents

Declaration ..... IV

Acknowledgements ..... V

Table of Contents ..... VII

List of Figures ..... X

List of Tables ..... XII

List of Equations ..... XIV

Abstract ..... 1

1 Introduction ..... 3

1.1 Aims of This Study ..... 3

1.2 Educational Relevance of This Study ..... 3

1.3 Structure of This Thesis ..... 5

2 Literature Review ..... 7

2.1 Definitions and conceptualisations ..... 8

2.1.1 Physical Activity ..... 8

2.1.2 Physical Education ..... 9

2.1.3 Exercise Intensity ..... 10

2.1.4 Exercise Complexity ..... 11

2.1.5 Behavioural Engagement ..... 12

2.1.6 Emotion States in the Classroom ..... 17

2.1.7 Dose-Response Relationship ..... 19

2.2 Theories of Physical Activity and Cognition ..... 22

2.2.1 Neurobiological Mechanisms ..... 23

2.2.2 Psychosocial Mechanisms ..... 27

2.2.3 Behavioural Mechanisms ..... 28

2.3 Research Evidence on Physical Activity and Cognition ..... 30

2.3.1 Chronic Physical Activity and Cognition ..... 31

2.3.2 Acute Physical Activity and Cognition ..... 33

2.3.3 Physical Activity and Behavioural Engagement ..... 35
2.3.4 Cognitively Engaging Physical Activity Breaks ..... 38
2.3.5 Physical Education Interventions ..... 40
2.3.6 Physical Activity, Emotion States, and Behavioural Engagement ..... 42
2.3.7 The Role of Individual Characteristics in the Physical Activity – Cognition Relationship ..... 45
2.4 Physical Education and Physical Activity in Primary Schools in the United Kingdom ..... 49
2.5 Gaps in the Current Literature ..... 50
2.6 Research Questions and Hypotheses ..... 53
3 Methodology ..... 56
3.1 Research Design ..... 58
3.2 Intervention Protocol ..... 59
3.3 Variables ..... 62
3.3.1 Independent Variable: Physical Activity ..... 63
3.3.2 Dependent Variable: Behavioural Engagement ..... 70
3.3.3 Mediation Variables: Emotions ..... 71
3.3.4 Control Variables ..... 73
3.4 Data Collection Instruments ..... 75
3.4.1 Observations ..... 75
3.4.2 Questionnaires ..... 80
3.5 Sampling Strategy ..... 82
3.5.1 Sample Characteristics ..... 83
3.6 Pilot Studies ..... 85
3.7 Reliability ..... 86
3.8 Validity ..... 90
3.9 Ethics ..... 91
3.10 Analysis ..... 92
3.10.1 Data Cleaning and Variable Computation ..... 92
3.10.2 Merging Data from Various Sources ..... 95
3.10.3 Analytic Strategies ..... 96
3.10.4 Application of Analytic Strategies ..... 101

3.11 Concluding remarks ..... 112
4 Results ..... 114
4.1 Sample ..... 114
4.1.1 Sample Descriptives ..... 114
4.2 Intervention Fidelity ..... 115
4.2.1 Physical Education Lesson Accelerometery Descriptives ..... 116
4.2.2 Intervention Fidelity: Accelerometery ..... 117
4.3 Physical Activity and Behavioural Engagement ..... 121
4.3.1 Observations Descriptives ..... 121
4.3.2 Multilevel Models of Physical Activity and Behavioural Engagement ..... 131
4.4 Physical Activity, Emotion States, and Behavioural Engagement ..... 167
4.4.1 Emotion State Descriptives ..... 167
4.4.2 Mediation Models of Physical Activity, Emotion State, and Behavioural Engagement ..... 168
4.5 Concluding Remarks ..... 186
5 Discussion ..... 188
5.1 Research Question 1 ..... 188
5.2 Research Question 2 ..... 192
5.3 Research Question 3 ..... 195
5.4 Research Question 4 ..... 197
5.4.1 Pre-PE On-Task Behaviour ..... 198
5.4.2 Attainment ..... 198
5.4.3 Sports Participation ..... 199
5.4.4 Gender ..... 200
5.5 Research Question 5 ..... 201
5.6 Implications for Practice ..... 204
5.7 Implications for Policy ..... 206
5.8 Strengths and Limitations ..... 210
5.8.1 Strengths ..... 210
5.8.2 Limitations ..... 211
5.8.3 Recommendations for Future Research ..... 212

6 Conclusion ..... 214

References ..... 215

## List of Figures

Figure 1. Arousal and valence of emotions in adults. Based on Inkinen et al. (2013). ..... 18

Figure 2. Three different dose-response relationships. A. Linear, B. Threshold, C. Inverted-U. ..... 19

Figure 3. Curve shifting as a result of task type (adapted from Schmidt & Wrisberg, 2004)..... 21

Figure 4. Proposed conceptual model of PA and mental and cognitive health (Lubans et al., 2016). 22

Figure 5. The human brain, with divisions between forebrain, midbrain, and hindbrain indicated (Romero, 2006:7). ..... 26

Figure 6. The four lobes of the cerebral hemispheres (Romero, 2006:103)..... 26

Figure 7. Proposed conceptual model of the pathways by which PA benefits academic achievement. ..... 52

Figure 8. The hypothesised effects of skill complexity (A) and exercise intensity (B) on subsequent behavioural engagement. ..... 54

Figure 9. The hypothesised interaction effect of skill complexity and exercise intensity. ..... 54

Figure 10. Content of the intervention lessons ..... 60

Figure 11. Observations app interface..... 79

Figure 12. Questionnaire app interface. ..... 81

Figure 13. Multilevel structure of observation data. ..... 96

Figure 14. The structural and measurement parts of SEM. ..... 99

Figure 15. A. Direct effect of X on Y. B. Effect of X on Y, mediated by M (Preacher, Rucker and Hayes, 2007: 188). ..... 100

Figure 16. Hypothesised factor structure. ..... 106

Figure 17. Simplified model with cross-lags. ..... 107

Figure 18. Simplified full SEM model. ..... 108

Figure 19. Hypothesised mediation effect of emotional state on behavioural engagement ..... 109

Figure 20. Sample mean proportion and 95% confidence interval of PE lessons spent at MVPA. ..... 118

Figure 21. Breakdown of moderate and vigorous PA by lesson condition. ..... 119

Figure 22. Mean and 95% confidence interval of mean ENMO during PE lessons by lesson intensity. ..... 120

Figure 23. Mean and 95% confidence interval of mean ENMO during PE lessons by PE condition. . 120

Figure 24. Model 1.4: Mean and 95% confidence interval of on-task behaviour pre- and post-PE. . 141

Figure 25. Random intercept-slope covariance for a random sub-sample of participants. ..... 142

Figure 26. Model 2: Mean and 95% confidence interval of on-task behaviour pre- and post-PE by complexity. ..... 144

Figure 27. Model 2: Effect size of PE lesson complexity. ..... 144

Figure 28. Model 3: Mean and 95% confidence interval of on-task behaviour pre- and post-PE by intensity. ..... 146

Figure 29. Model 3: Effect size of PE lesson intensity. ..... 146

Figure 30. Model 4: Mean and 95% confidence interval of on-task behaviour pre- and post-PE by complexity-by-intensity. ..... 149

Figure 31. Model 4: Effect size of PE lesson complexity-by-intensity. ..... 149

Figure 32. Effect size of PE lesson intensity-by-complexity for the whole sample and those below the 33rd percentile on-task. ..... 154

Figure 33. Effect of PE lesson intensity-by-complexity-by-attainment. ..... 159

Figure 34. Effect of PE lesson intensity-by-complexity-by-sports participation. ..... 163

Figure 35. Effect of PE lesson intensity-by-complexity-by-gender. ..... 166

Figure 36. Invariance testing. ..... 176

Figure 37. Autoregression model including PE lesson (standardised results). ..... 178

Figure 38. Correlation at lesson level (standardised results). ..... 179

Figure 39. Regression of emotion state at lesson level (standardised results). ..... 180

Figure 40. Regression of emotions at lesson level, with cross-lags (standardised results). ..... 181

Figure 41. Regression with cross-lags and mediation paths (standardised results). ..... 182

Figure 42. PE lesson content predicting PE emotions (standardised results). ..... 183

Figure 43. PE lesson content and enjoyment predicting emotions (standardised results). ..... 184

Figure 44. PE lesson content, emotions, and task-behaviour (standardised results). ..... 185

## List of Tables

Table 1. Example data for calculation of chance agreement of raters. ..... 88

Table 2. Data structure in long format (raw data, used for MLM). ..... 92

Table 3. Data structure in wide format (aggregated data, used for SEM). ..... 93

Table 4. Overview of physical activity data analysis strategy. ..... 110

Table 5. Sample descriptives ..... 115

Table 6. Duration at intensity by intervention lesson ..... 116

Table 7. Pairwise comparisons of MVPA duration during PE lessons ..... 117

Table 8. 3-by-2 ANOVA results. ..... 118

Table 9. Factorial ANOVA of mean ENMO during PE. ..... 119

Table 10. Distribution of observations across intervention conditions and year groups ..... 122

Table 11. Behaviour descriptives whole sample and by ID. ..... 123

Table 12. Distribution of task-type in pre- and post-PE lessons for the sub-sample with attainment scores. ..... 124

Table 13. Self-report data descriptives for observed sample with attainment scores. ..... 125

Table 14. Ratings of tiredness during classroom lessons ..... 126

Table 15. Task behaviour by task enjoyment. ..... 127

Table 16. Task behaviour by task type. ..... 127

Table 17. Crosstabulation of PE lesson design and PRE. ..... 128

Table 18. Task behaviour by year group, gender, and BMI. ..... 129

Table 19. Task behaviour by attainment. ..... 129

Table 20. Model 1.1-1.3: fixed and random parameters. ..... 133

Table 21. Effect size of PrePost with stepwise addition of control variables. ..... 134

Table 22. Stepwise addition of control variables in model 1.4. ..... 135

Table 23. Model 2: PrePost-by-Complexity: parameter estimates ..... 143

Table 24. Model 3: PrePost-by-Intensity: parameter estimates. ..... 145

Table 25. Model 4: Interaction of intensity-by-complexity - parameter estimates. ..... 148

Table 26. Model 5.1: Intensity-by-complexity-by-pre-PE on-task - parameter estimates. ..... 151

Table 27. Chi^2 for each intervention condition, split by teacher-rated attainment. ..... 154

Table 28. Model 5.2: Intensity-by-complexity-by-attainment - parameter estimates. ..... 155

Table 29. Model 5.3: Intensity-by-complexity-by-sports participation - parameter estimates. ..... 160

Table 30. Model 5.4: Intensity-by-complexity-by-gender- parameter estimates. ..... 164

Table 31. Percentage high emotion ratings in classroom and PE reports ..... 167

Table 32. Missing data in self-reports. ..... 168

Table 33. Overview of SEM goodness of fit. ..... 169

Table 34. Goodness of fit indices for 8-emotion EFA. ..... 170

Table 35. Standardised factor loadings for 3-factor model with 8 emotions. ..... 171

Table 36. Goodness of fit indices for 7-emotion EFA excluding ‘excited’. ..... 171

Table 37. Standardised factor loadings for 2- and 3-factor model with 7 emotions excluding ‘excited’. ..... 172

Table 38. Goodness of fit indices for 7-emotion EFA, excluding ‘bored’. ..... 172

Table 39. Standardised factor loadings for 2- and 3-factor model with 7 emotions excluding 'bored'. ..... 173

Table 40. Goodness of fit indices for 6-emotion EFA. ..... 173

Table 41. Standardised factor loadings for 2- and 3-factor model with 6 emotions. ..... 173

Table 42. Goodness of fit indices for 3-level CFA model. ..... 174

Table 43. Standardised factor loadings for multilevel CFA. ..... 175

Table 44. Goodness of fit indices for invariance models. ..... 176

Table 45. Positive emotion correlations across the day. ..... 177

Table 46. Negative emotion correlations across the day. ..... 177

Table 47. Goodness of fit indices for autoregressive models. ..... 178

Table 48. Goodness of fit indices for autoregression model with and without cross-lags. ..... 181

Table 49. Goodness of fit indices for emotion state model with and without predictors. ..... 182

Table 50. Goodness of fit indices for model with and without behaviour outcomes. ..... 185

## List of Equations

Equation 1. Agreement expected by chance (McHugh, 2012). ..... 88

Equation 2. Actual agreement as a raw percentage (McHugh, 2012). ..... 88

Equation 3. Cohen's Kappa (McHugh, 2012). ..... 89

Equation 4. Standard error of Cohen's Kappa (McHugh, 2012). ..... 89

Equation 5. 95% Confidence Interval for Cohen's Kappa (McHugh, 2012). ..... 89

Equation 6. Null Model. ..... 98

Equation 7. Intra Class Correlation coefficient. ..... 98

Equation 8. Regression without mediator. ..... 100

Equation 9. Regression with mediator. ..... 101

Equation 10. Mediation effect. ..... 101

Equation 11. Model 1.1: Fixed effects. ..... 102

Equation 12. Model 1.2: Random intercepts. ..... 102

Equation 13. Model 1.3: Random intercepts and slope. ..... 103

Equation 14. Model 1.4: PrePost with control variables. ..... 103

Equation 15. Model 2 (RQ1). ..... 103

Equation 16. Model 3 (RQ2). ..... 103

Equation 17. Model 4 (RQ3). ..... 104

Equation 18. Model with interaction effects for gender. ..... 104

Equation 19.Direct effect of acute PA on post-PE on-task behaviour. ..... 108

Equation 20. Mediated effect of acute PA on post-PE on-task behaviour. ..... 109

Equation 21. Mediated effect of acute PA on on-task behaviour via emotion state. ..... 109

Equation 22. Direct effect of PE enjoyment on post-PE on-task behaviour. ..... 109

Equation 23. Mediated effect of PE enjoyment on post-PE on-task behaviour. ..... 109

Equation 24. Mediated effect of PE enjoyment on on-task behaviour via emotion state. ..... 109

Equation 25. Participant level ICC. ..... 131

Equation 26. Week level ICC. ..... 131

Equation 27. Model 1.1: Pre/Post PE behaviour – fixed effect. ..... 132

Equation 28. Model 1.2: Pre/Post PE behaviour – random intercept. ..... 132

Equation 29. Model 1.3: Pre/Post PE behaviour – random intercept and random slope. ..... 132

Equation 30. Model 1.4: Pre/Post PE behaviour – random intercept and random slope with all control variables. ..... 141

Equation 31. Model 2: PrePost-by-Complexity. ..... 143

Equation 32. Model 3: PrePost-by-Intensity. ..... 145

Equation 33. Model 4: PrePost-by-Complexity-by-Intensity. ..... 147

Equation 34. Model 5.1. Intensity-by-complexity-by-pre-PE on-task behaviour..... 153

Equation 35. Model 5.2. Intensity-by-complexity-by-attainment. ..... 158

Equation 36. Model 5.3. Intensity-by-complexity-by-sports participation..... 162

Equation 37. Model 5.4. Intensity-by-complexity-by-gender. ..... 166

# Abstract

*Objectives*: To investigate the effect of (i) Physical Education (PE) lesson intensity and (ii) skill complexity, (iii) their interaction on students' on-task behaviour in the classroom, (iv) the role of individual differences in the effect of PE lesson intensity and complexity on on-task behaviour, and (v) the mediation of this effect by emotion states.

*Method*: Participants were children (N=101, age 7-10) recruited from four primary schools in Oxfordshire, United Kingdom. The intervention consisted of manipulating the aerobic intensity (low/medium/high) and skill complexity (low/high) of PE lessons. Children participated in all six conditions of the experiment: low intensity/low complexity (flexibility), medium intensity/low complexity (health related exercise), high intensity/low complexity (sprinting games), low intensity/high complexity (bi-lateral ball skills), medium intensity/high complexity (ball games), high intensity/high complexity (aerobics). Children's behaviour in the classroom was observed every 30 seconds for 25 minutes before and after each PE lesson and rated as on-task or off-task. Participants completed self-reports on tablets three times per classroom lesson and once at the end of each PE lesson. Behaviour was analysed with multilevel models, as well as individual differences in participants' response to physical activity (PA). Mediation through emotion states was investigated with structural equation models.

*Results*: No main effect of complexity on on-task behaviour was observed. No main effect of intensity on children's on-task behaviour was found, although on-task behaviour after high-intensity PE lessons was significantly higher than before PE ($\chi^2_{high\ intensity}$=4.834, p<0.05), the effect size did not reach the cut-off for a small effect (d=0.19, OR=1.41). There was an interaction effect of intensity and complexity, and high intensity/high complexity was the only condition to reach significance ($\chi^2_{high-high}$=12.197, p<0.001) with an effect size greater than d=0.2 (d=0.33, OR=1.82). There were individual differences in participants' response to PA; those with the lowest attainment level and those

who were least on-task before PE had the strongest response to PA, and girls had a more positive response than boys. There was no significant mediation of the relationship between PE lesson content and subsequent on-task behaviour by emotion states or PE lesson enjoyment.

*Conclusions*: These findings suggest that participation in PE lessons can improve children's on-task behaviour in the classroom, in particular for those students with low levels of on-task behaviour. PE lessons involving high complexity and high intensity appear to have the greatest benefits for students' behaviour in the classroom.

Word count: 57,653

# 1 Introduction

## 1.1 Aims of This Study

Results from recent studies indicate that it might be possible to make therapeutic use of exercise to improve cognitive functioning. However, knowledge of the full range of effects of exercise on cognitive functioning has not yet been achieved, and current studies continue to elucidate aspects of this complex relationship. Whilst researchers are uncovering the positive effects of physical activity (PA) on cognition, only half of British seven-year-olds achieve the guideline amount of physical activity (PA) (Griffiths *et al.*, 2013), and PA opportunities in schools are declining (Harris, Cale and Musson, 2012). Considering these two factors, research into the relationship between PA in school-aged children and academic achievement and its precursors, such as levels of student engagement in the classroom, is vital. This study aims to provide a better understanding of how Physical Education (PE) lesson content can benefit primary school children's learning by assessing the dose-response relationship between PE lesson intensity and complexity and behavioural engagement, if there is an interaction effect of intensity and complexity, and what the role of individual characteristics is in these relationships. Finally, it was investigated if any of the aforementioned relationships are mediated by emotion states.

## 1.2 Educational Relevance of This Study

Teachers have noticed a decrease in pupils' engagement (Budde *et al.*, 2008; Yu *et al.*, 2006). This development is cause for concern, as inattentiveness has been linked to lower attainment (Roorda *et al.*, 2011; Pellegrini and Davis, 1993). Pellegrini and Davis (1993) observed third grade children during lessons, and correlated their verbal and non-verbal scores on standardised tests with their fidgetiness

during lessons (d=-0.45)[1], and their concentration (d=0.52). Alongside decreasing engagement, a decrease in children's activity levels has been noted over the past decade, and around half of all seven-year-old children do not meet the guidelines for PA (Griffiths *et al.*, 2013). It has been argued that increased achievement pressure combined with the general belief that PE is of lower educational value than lessons in formal academic topics have led to a decrease in PA levels in schools (Chaddock *et al.*, 2011). Yet, previous studies into the relationship between PA during the school day and academic performance have found that increasing PA does not correlate with declining performance, and in many cases is related to improved cognitive performance (Pellegrini and Smith, 1993; Chaddock *et al.*, 2011). Coe *et al.* (2006) found that pupils who achieved the US Healthy People 2010 guidelines for PA performed better in school assessments compared to those who were less active ($\chi^2$=10.1, p<0.01, d=0.45). Additionally, a range of cognitive skills and functions have been found to be improved by exercise, e.g. vocabulary recall performance ($F_{(2,66)}$=4.82, p=0.03, d=0.59) (Salis, 2013), and selective attention (t=2.34, p=.02, d=0.50) (Alves *et al.*, 2014).

The link between PA and cognitive performance continues to be a much investigated topic both in the UK and outside, as recent publications by for example Valentini *et al.* (2017), Westfall *et al.* (2017), and Haapala *et al.* (2016) demonstrate. However, no such work has been published with regards to the effect of acute exercise intensity and complexity on subsequent behavioural engagement in the classroom. Although several recent experimental studies have found positive effects of (specific forms of) PA on cognitive functioning of adults (e.g. Hoetting *et al.*, 2016; Weinberg *et al.*, 2014; Schmidt-Kassow *et al.*, 2013; Schmidt-Kassow *et al.*, 2014), these were conducted in laboratories, and their findings are yet to be applied to an educational setting by modifying and utilising the content of PE lessons in order to improve behavioural engagement.

Pupils do not increase their cognitive functioning by simply participating in any given form of PA. This is not a new concept; the specific content and delivery of PE lessons have previously been

[1] Where possible, effect sizes throughout this thesis have been calculated as Cohen's d by the author, using equations from Field (2013), Lenhard and Lenhard (2016), and Borenstein *et al.* (2009). For more information see appendix A.

identified by Bailey (2006) and Lubans *et al.* (2017) as important factors in the extent to which pupils benefit from them. Thus, a thorough understanding of PE is important for teachers. However, in most primary schools in the UK, PE is delivered by classroom teachers. These are not the subject specialists found in secondary schools, and research into primary school teachers' perceptions of PE has demonstrated that PE is not used to its full potential due to teachers' lack of understanding and confidence (Morgan and Hansen, 2008b).

Until recently, UK primary schools outside the independent sector have received a PE and Sport Premium from the Secretary of State for Education (2016a). This funding, which was first introduced in 2013 (Ofsted, 2014), and stood at £8000 plus £5 per pupil in the 2016-17 academic year, was aimed at supporting primary schools to encourage their pupils to have active and healthy lifestyles, and to make sustainable improvements to their current provision (2016a). This government initiative, part of the London 2012 Olympic legacy, aimed to counter the current trend of decreasing PA levels in children and young people, but was set to come to an end in 2016 (Ofsted, 2014). It was extended into the 2016-17 academic year (2016b), but no commitment to further extension was made by Government and the funding has now ceased.

Through the PE and Sport Premium, teachers in many schools have had the assistance of specialist coaches in delivering high quality PE and sport for the past five years. Upon the cessation of this funding, they will now need to take the reins themselves. This creates an optimal opportunity to supply teachers with information on the newly discovered benefits of PE for cognitive function and behavioural engagement.

## 1.3 Structure of This Thesis

This thesis will commence with an overview of the current research field, provided in chapter 2, the literature review. This chapter will first of all introduce and explain the applicable concepts, and relate these to the current study. Next, the relevant literature in the field of PA and cognition will be

presented, and research evidence will be provided for the theories which underpin my study and hypotheses. Moreover, the context of this study, primary school PE lessons in the United Kingdom, will be described. In chapter 3, the research design and protocols are provided along with the rationale for the research method and instrument selection and development. To conclude the methodology chapter, the analysis strategies and data management procedures are explained. Chapter 4 presents the analysis results and the answers to each of the research questions, which are discussed in detail and related to the existing literature in chapter 5. Chapter 5 furthermore identifies the implications of this study for practice and policy. To conclude, it highlights the strengths and limitations of this study, and presents suggestions for future research.

# 2 Literature Review

This chapter presents an overview of the existing literature on the topics of behavioural engagement in the classroom, acute PA, and their relationship to one another. It commences by defining the concepts involved in this study, including the PA, exercise intensity and complexity, behavioural engagement, emotion states, and the dose-response relationship. Then, the current ideas and theories about the relationship between acute and chronic PA and measures of cognition, including on-task behaviour, are introduced. Research evidence of the links between PA, cognition, and emotion states is discussed in relation to the existing theories. Next, the context in which this study was conducted – PA in primary schools in the United Kingdom - is described. Finally, gaps in the current literature are highlighted and the research questions and hypotheses for this study are provided.

The theories on PA and cognition have foundations in the fields of neurobiology, psychology, education, and sports and exercise sciences, and research findings are diverse. Yet it is widely accepted across the literature that a positive correlation exists between participation in acute PA and fitness on the one hand, and cognitive functions and academic achievement on the other hand. In a meta-analysis of 59 studies, Fedewa and Ahn (2011) found that there was a significant and positive impact of all PA programmes on children's cognitive outcomes and academic performance (d=0.35). However, the exact nature of this relationship has not yet been fully elucidated, despite a large volume of research being conducted in this field. Countless parameters exist for the type, duration, intensity, and timing of the exercise, and as least as many for the cognitive task. Added to this are individual factors such as participants' age and fitness level. Altogether, the relationship between PA and cognition is a highly complex one.

Results from studies based on tests of cognitive ability or academic grades give a global overview of the magnitude of the effect of PA on participants' cognitive functioning, but fail to clarify exactly

how and why PA affects it. Based on research findings, several theories have been developed about the relationship between PA and cognition, including neurobiological, emotional, social, and behavioural mechanisms. In this chapter, these theories will be discussed and related to primary school children's classroom-based learning and PE lessons.

## 2.1 Definitions and conceptualisations

The first section of this chapter will introduce the concepts involved in this study. First, the concepts regarding the independent variable (IV), PA, will be explained. These are followed by the concepts related to the dependent variable (DV), behavioural engagement, and the mediation variable, emotional state. Then, the mechanism of the dose-response relationship, which links the IV and DV, will be described. Last, the influence of individual differences on the dose-response relationship will be explored.

### 2.1.1 Physical Activity

As stated by Cumming and Riddoch (2008), PA is a behaviour, and a highly complex one. It can be defined as behaviour which involves movement of (part of) the body (Schmidt-Kassow *et al.*, 2014). There is a vast range of behaviours that qualify under this definition. There is great variation in the nature of the PA, such as type, duration, and intensity (both absolute and relative to the individual's maximal capacity), and there are also behavioural aspect to consider such as the frequency of participation, the energy expenditure or metabolic cost, whether the activity was voluntary or unavoidable (such as active transport when walking or cycling to school or work), and the social and cultural context in which the activity takes place (Cumming and Riddoch, 2008). A variety of individual characteristics, such as body type, age or maturation stage, skill level and fitness level, in turn affect the behaviour (Schmidt and Wrisberg, 2004). In this study, the term PA will refer to any activity

involving purposeful bodily movement, other than sedentary behaviours (such as arm movements for writing, whilst sitting at a desk).

It is important to differentiate between acute and chronic PA, as the effects on cognitive function are different. Acute PA involves participation in an isolated PA event, also called an exercise bout, and the effects related to this type of PA are transient (Chang *et al.*, 2012). Research has typically investigated effects immediately after participation, with follow-ups of up to one week later (e.g. Lundbye-Jensen *et al.*, 2017; Frith, Sng and Loprinzi, 2017). Chronic PA requires regular participation over an extended period of time, and effects of this type of PA are more structural and lasting (Chaddock *et al.*, 2011; Cumming and Riddoch, 2008). Research involving chronic PA often includes measures of fitness as a dependent variable (DV), along with cognitive measures (e.g. Aadland *et al.*, 2017; Haapala *et al.*, 2016; Chaddock *et al.*, 2011). This chapter will give an overview of the research landscape involving both types of PA, before focusing on the acute effects.

### 2.1.2 Physical Education

Physical Education is PA which takes place in a particular format and context. It is a school lesson, and therefore has a specific learning goal in line with the school or country's curriculum plan. In the United Kingdom, the Primary National Curriculum states the aims for PE as:

"to ensure that all pupils:

- develop competence to excel in a broad range of physical activities
- are physically active for sustained periods of time
- engage in competitive sports and activities
- lead healthy, active lives." (Department for Education, 2013: 198).

The PA in PE lessons is (generally) teacher-directed, and of a predetermined and extended duration; between 30 and 60 minutes in most schools. The variety and range of activities children are presented with in their PE lessons are dependent on the teacher's level of experience and knowledge within the subject.

In the UK, the content and structure of PE lessons can vary greatly at primary school level. Very few schools employ specialist PE teachers, and generally the classroom teacher is responsible for delivering the PE lessons. Differences in the teachers' own experience of PE lessons when they were in school, their general beliefs regarding the importance of PE and sports, their confidence in teaching the subject, and how much time was dedicated to PE during their teacher training are among the factors which contribute to this variation in their PE delivery style and the range of content they teach (Freak and Miller, 2017; Harris *et al.*, 2012; Morgan and Hansen, 2008a). Additionally, there are school-level factors such as timetable demands, and the availability of equipment and space (Tsangaridou, 2017; Rainer *et al.*, 2012). Finally, there is no government directed minimum frequency or duration for the provision of PE in UK primary schools; all children must be provided with regular PE lessons, but that is as far as the legislation goes (Department for Education, 2013). It is recommended schools aim for two hours of PE a week, but this is only advice. Weekly duration of PE provision is one of the four areas scored in the 'Healthy Schools' rating scheme by the Department for Education (Department for Education, 2019). Schools much provide all year groups with a minimum of two hours of PE a week to be eligible for the silver or gold level 'Healthy Schools' rating. However, participation in the rating scheme is voluntary (ibid.).

### 2.1.3 Exercise Intensity

The aerobic intensity of PA is related to the demand placed on the cardiovascular system, and in laboratories this is often measured by the volume of oxygen breathed in during exercise ($VO_2$), or by heart rate, expressed as a percentage of the individual's $VO_2$ or heart rate at maximal exercise (Fawkner and Armstrong, 2008; Rowland, 2008). Heart rate can be easily recorded with non-invasive methods out in the field also, but to get a precise and individualised measure of exercise intensity you would need to first assess the participant's individual maximum heart rate in a progressive or maximal effort exercise test. The same applies to $VO_2$; without the maximum scores as a reference, the raw

data can only approximate the exercise intensity (Rowland, 2008; Fawkner and Armstrong, 2008; Armstrong and Fawkner, 2008).

Another non-invasive method used for estimating exercise intensity is with accelerometers. These record the amount of movement which occurs at the site that the accelerometer is worn. Common sites to wear these devices are the hip, ankle, and wrist (McLellan, Arthur and Buchan, 2018; Activinsights, 2012). Cut points can be applied to the level of movement recorded to classify the exercise as sedentary, light, moderate, or vigorous (Duncan *et al.*, 2016). Most established devices used in a research setting have been validated against heart rate or VO2 data (Phillips, Parfitt and Rowlands, 2013).

This study used accelerometers to record movement data and approximate the intensity of the PA participated in during the intervention period. More information on the devices used can be found in section 3.3.1, and validation information can be found in appendix B.

### 2.1.4 Exercise Complexity

The concept of exercise complexity has recently received a great deal of attention. It has been recognised that the demands placed on the body by exercise do not only vary in how they stress the cardiovascular system, but also by how much they strain the motor areas of the brain. Exercise complexity can be achieved in two ways; by increasing the complexity of the movements themselves, or by overlapping cognitive demands with the PA (Budde *et al.*, 2008; Schmidt *et al.*, 2015b).

Simple, routine movements ingrained in the individual's daily movement patterns, provide a very different workout from movements which are complex, novel and not yet ingrained, even if they are performed at the same exercise intensity. The addition of elements of balance, the coordination of multiple limbs at the same time, performing exercises with the non-dominant hand or foot, or the pressure of needing to keep in time can intentionally increase the demands of the activity on the brain, with coordinative exercise influencing several neuro-behavioural systems (Budde *et al.*, 2008).

Cognitive demands overlapped on exercise naturally occur in many sports and activities, for example rules which are applied to games, and sequences of movements which need to be remembered in gymnastics or dance. The cognitive demand of an activity can be purposely increased by frequently changing the rules, or by changing the order or adding elements to sequences (Schmidt *et al.*, 2015b). Another way to increase cognitive demand is by asking participants to perform cognitive tasks whilst exercising. This can be easily achieved in a school setting by adding curriculum content from other subjects to the PE lesson.

In this study, cognitive demand during the PE lesson was manipulated through the use of complex and novel movements in the high intensity lessons, and by deliberately increasing the naturally occurring cognitive demands of the exercises performed with high-paced changes between activities. In low complexity lessons, simple, routine movements were used, and rules for activities were kept to a minimum.

### 2.1.5 Behavioural Engagement

In studies of classroom behaviour, the terms 'on-task', 'engagement', and 'attention' are used without harmony of the precise operationalisation, and the definitions and concepts vary between studies, depending on the research question(s) addressed. Common measures are time spent on-task, or off-task (Grieco *et al.*, 2016; Szabo-Reed *et al.*, 2017; Godwin *et al.*, 2013). In order to be, and stay, on-task, students need to direct their attention to the correct target and keep it there, a process also known as attentional allocation. Three phases of attentional allocation occur in the brain: alerting, orienting, and executive attention. The execution of these three phases leads to attentional allocation based on the detection, prioritisation, and selection of stimuli, as well as inhibition of competing stimuli and incorrect dominant responses.

## Alerting

Firstly, for attention to be directed at a target, this target (a stimulus) needs to be detected (Broadbent, 1953). This is a process dependent on alertness, in which stimuli are prioritised based upon their physical intensity, biological importance or motivational valence, and novelty level (Broadbent, 1953; Frankmann and Adams, 1962). In the mid-20th century, alertness in adults was found to decrease during a monotonous, prolonged task (Frankmann and Adams, 1962), most strongly so after the first half hour, with the decline from the 2nd to the 4th half hour non-significant (Mackworth, 1948). The rationale provided for this phenomenon was the effect of novelty level of a stimulus (Frankmann and Adams, 1962). At the start of a task, the novelty level of the situation is high, leading to correspondingly high alertness levels (Mackworth, 1965). As habituation takes place due to stimulus repetition or absence, the level of brain activity decreases, in similar ways as during sleep (ibid.). Habituation occurs sooner in increasingly monotonous environments (ibid.).

The early vigilance studies have been backed by the results of contemporary research into the role of various neurochemicals during alerting (Robbins, 2002; Nieuwenhuis, Aston-Jones and Cohen, 2005). Alertness naturally fluctuates through the day and is affected by the levels of the neuromodulator norepinephrine (NE) (Posner, Rueda and Kanske, 2007). As a result of NE presence, synapses in the affected brain regions become more sensitive to both excitatory and inhibitory neurotransmitters, leading to enhanced cognitive functioning (Nieuwenhuis *et al.*, 2005; Robbins, 2002). Studies of alertness during cognitive tasks found that participants who were more alert were able to complete tasks with decreased reaction times (Posner *et al.*, 2007). Alertness is also crucial for optimal completion of concurrent tasks (ibid.).

All behavioural responses to stimuli are preceded by a spike in NE expression (Nieuwenhuis *et al.*, 2005). Phasic expression is the result of a task-relevant stimulus (target), or a task-irrelevant stimulus (distractor) (ibid.), which is either physically strong, motivationally valent, or novel (Nieuwenhuis *et al.*, 2005; Broadbent, 1953). In the case of novel stimuli, the NE response habituates with repeated occurrence of the stimulus (Nieuwenhuis *et al.*, 2005). Stimuli which are motivationally significant or

emotionally loaded, whether positive or negative, are more likely to elicit spiked NE expression (ibid.). During medium baseline NE expression, with spiked expression as a result of target stimuli, restricted attention is paid to those stimuli which are task-relevant, leading to goal-directed behaviour (ibid.). NE expression therefore functions as an attentional filter, determining when attention must be focused (at the occurrence of target stimuli) rather than exploratory or sleep-like (ibid.).

In short, alerting is the process by which NE expression is regulated, and stimuli are detected (or not). It is a prerequisite for attentional control. Once a stimulus has been detected, it can be explored or ignored. At this point orienting – the process of selecting the strongest, most relevant stimuli – comes in.

### Orienting

Orienting depends on the interplay of NE and the neurotransmitter acetylcholine (ACh). Together, these regulate the individual's susceptibility to distraction, the likelihood of involuntary shifts of attention, and the accuracy of their stimulus response (the correctness of the stimulus they have selected to attend to).

NE plays a role in orienting as well as alerting processes; lesions of rats' prefrontal cortex (depleting NE) showed a disproportionately greater decline in stimulus response accuracy when distractors were present during task execution (Robbins, 2002). These results indicate that NE increases innervation of the neuronal structures leading to the preservation of attentional selectivity in high-arousal situations to optimise accuracy (ibid.). This effect of NE is related to its effect on neurotransmission. Orienting depends on cholinergic neurotransmission in the forebrain (Davidson and Morrocco, 2000; Robbins, 2002). NE makes the synapses in the cortical areas more responsive to cholinergic neurotransmitters (Nieuwenhuis *et al.*, 2005). As well as the responsiveness of synapses, varying levels of the cholinergic neurotransmitter ACh affect the brain's ability to direct and sustain attention (Davidson and Morrocco, 2000). Robbins (2002) observed that ACh is naturally at a lower level in novel circumstances compared to familiar, or pre-trained, circumstances. A decrease in ACh, or a blocking of the receptors,

is manifested in slower RT, increased error rates, and increased distractibility (Davidson and Morrocco, 2000; Robbins, 2002). NE is therefore required in higher doses in novel environments to counteract distractibility due to low levels of ACh. Thus, NE release not only increases alertness, but also improves the brain's responsiveness to ACh, and facilitates sustained attention at the orienting stage.

In conclusion, orienting is the stage at which the detected stimuli are evaluated, and the decision is made which stimulus to respond to, and which to ignore. Orienting only differentiates between stimuli which are dominant or subordinate. When there is a conflict between stimuli which are of equal importance, or if a subordinate stimulus or response must be prioritised over a dominant one, the executive attention system takes over.

### Executive Attention

The brain makes decisions for action based on stimuli and prioritises response to certain stimuli whilst ignoring others (Berlyne, 1966; Frankmann and Adams, 1962). The choices made are dependent on contextual variables (Botvinick *et al.*, 2001). The executive attention system is activated when there is a conflict between stimuli or responses, requiring a sub-dominant response to be selected (ibid.). As part of executive attention, conflict monitoring occurs to control attention, and self-correction takes place when attention is paid to the incorrect stimulus or target (Posner *et al.*, 2007). Involuntary attentional shifts may be prevented by the process of inhibition (ibid.). With age, humans develop the neuronal structures in the frontal lobe, the brain area responsible for inhibition (Miyake *et al.*, 2000; Buck, Hillman and Castelli, 2008), and their susceptibility to visual distractors decreases (Godwin and Fisher, 2011).

Executive attention is the point at which executive function (EF) becomes involved in the attentional allocation process. Tests of executive attention present intentionally conflicting stimuli, such as the incongruent condition of the Stroop task or the Flanker task (Posner *et al.*, 2007). More details on frequently used tests of attention, inhibition, and concentration can be found in appendix C. Attentional allocation is also relevant in classroom-based learning situations; for example, if a

physically strong stimulus such as a loud noise occurs, yet pupils must maintain their focus on the task they are completing.

### Importance of Behavioural Engagement in the Classroom

The idea that attention and behavioural engagement are required for academic performance is not new. In recent decades, behavioural engagement and time-on-task have been highlighted by many authors as being crucial to academic achievement and success (e.g. Pellegrini and Holmes, 2006; Godwin and Fisher, 2011; Watson *et al.*, 2019). Fredricks, Blumenfeld and Paris (2004) describe how disruptive and inattentive students have been found to score lower in tests, and how this correlation holds across all ages, from early primary to the end of secondary school. More specifically, Hirvonen *et al.* (2010) found that on-task behaviour in preschool predicted later spelling achievement (grades 1-4) by 3-5%. It is clear that on-task behaviour is an important factor in academic achievement, and that increasing on-task behaviour is a potential contributor to improve students' academic outcome.

### Operationalisation of On-Task Behaviour

In this study, on-task behaviour is operationalised as 'displaying goal-directed and task-appropriate behaviours'. The 'goal-directed' element of the operationalisation accounts for the absence or irrelevant or distracted behaviour such as yawning, day-dreaming, or unnecessary movement or communication. The inclusion of 'task-appropriate' implies a dependence of the displayed behaviour on the task that is being performed. When children are asked to work quietly and independently on a writing task, on-task behaviour will look very different from when they are working in a group, and are asked to act out a story they have written. The observation protocol used reflects the dependence of behaviour on task type, and records task type alongside participants' behaviour so that it can be controlled for in analyses. A more detailed operationalisation of on-task behaviour can be found in sections 3.3.2 and 3.4.1, as well as appendix D.

### 2.1.6 Emotion States in the Classroom

Emotion states are thought to be crucial to the motivational processes behind a range of behaviours (Dunton *et al.*, 2014; Hamaker, Nesselroade and Molenaar, 2007). Emotions prepare us for reactions to important events by providing motivational and physiological energy, focussing attention, and prompting action-related intentions (Pekrun *et al.*, 2002). Thus, emotions can greatly affect students' thoughts, motivation, and behaviour, making them of interest for classroom behaviour. Emotion states fluctuate throughout the day; they can last from a few seconds to a few hours (Newton, 2013). These are not to be confused with moods, which are relatively much longer lasting, or personality traits, which are relatively fixed (ibid.). Besides being comparatively stable, traits are assumed to be mostly person-specific, whereas states are event- or situation-specific (Hamaker *et al.*, 2007). Variation in emotion states contributes to variations in behaviour, as the combination of trait and state emotions influences the likelihood a person reacts to a situation in a certain way (ibid.). For example, a person who generally has a positive outlook (trait), is likely to respond to the negative emotions induced by an unpleasant event (state) differently than a person who has a predominantly negative outlook.

As well as influencing pupils' psychological well-being, emotion states in the classroom affect cognitive, motivational, and regulatory processes involved in learning and achievement (Pekrun, 2006; Goetz *et al.*, 2006). Emotions are relevant to learning and behaviour, as they can be task-promoting or task-inhibiting, depending on their arousal value (activating or deactivating) and valence (positive or negative) (Newton, 2013; Inkinen *et al.*, 2013). Activating emotions are those which more readily lead to taking action and accompany an energised feeling (Newton, 2013). Examples include happiness, anger, and frustration. On the other hand, deactivating emotions are those which are associated with feeling de-energised, such as sadness and serenity (Inkinen *et al.*, 2013). Positive emotions are those emotions which are pleasant experiences, whereas negative emotions are unpleasant experiences (Rush and Hofer, 2014). The combination of arousal and valence of emotions create one's emotional state, or affect, as illustrated in Figure 1 (Inkinen *et al.*, 2013).

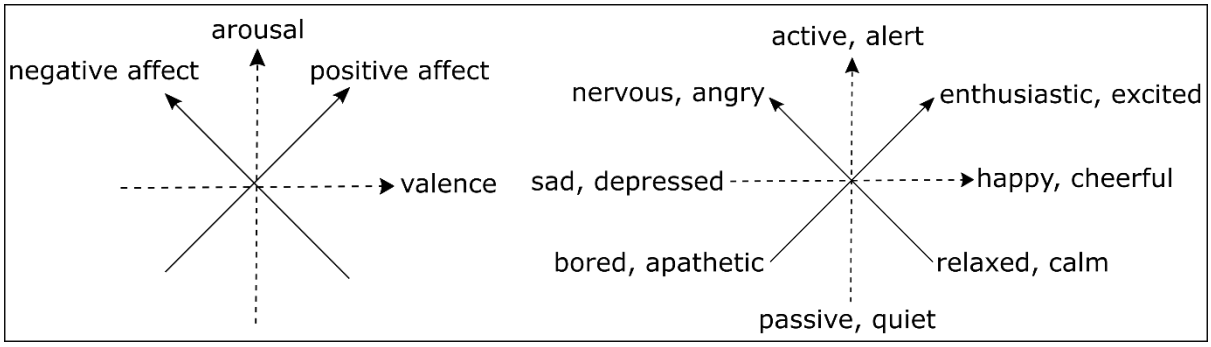



*Figure 1. Arousal and valence of emotions in adults. Based on Inkinen et al. (2013).*

Both positive and negative emotions can present a distraction from the task at hand by inducing task-irrelevant thoughts and leading to mind-wandering and disengagement (Pekrun, 2006; Scrimin and Mason, 2015). Generally, positive activating emotions such as enjoyment are task-promoting (Pekrun, 2006; Dunton *et al.*, 2014), but deactivating positive emotions, such as relaxation, are less straight-forward. These can lead to either task avoidance in extrinsically motivated pupils or increased task focus in intrinsically motivated individuals (Pekrun *et al.*, 2002). Negative deactivating emotions (e.g. sadness, boredom) tend to be detrimental to task focus, whereas negative activating emotions also present a more complex situation (Pekrun, 2006). Negative activating emotions can be detrimental due to their negative valence which reduces intrinsic motivation, or they could facilitate greater extrinsic motivation due to their activating nature stimulating the individual to take action and tackle the problems leading to the negative feelings (Pekrun *et al.*, 2002). An often-cited example of this type of emotion is anxiety, which can lead either to avoidance of the task that could lead to failure, or to increased task focus to deal with the task and avoid failure (e.g. Dunton *et al.*, 2014; Pekrun *et al.*, 2002). Consequently, emotions can affect engagement and task performance; making emotion states highly relevant to behavioural engagement.

As mentioned previously, emotion states are related to experiences such as events and people (Newton, 2013) and fluctuate throughout the day (Hamaker *et al.*, 2007). Thus, pupils' emotions at the start of a lesson may not be the same as at the end of said lesson, and the events and people encountered in one lesson may influence feelings during the next lesson. As emotions are influenced by events, PA opportunities during the school day may affect students' emotion states. Section 2.2.2

describes theories of how PA relates to emotion states, and section 2.3.6 presents research evidence on the links between PA, emotion states, and behavioural engagement. This study collected self-reports of pupils' emotional state at three points per classroom lesson, and at the end of their PE lessons. These ratings were used to investigate if emotional states are affected by PE lesson content, and whether emotional state has a mediating effect on behavioural engagement after PE lessons. The next section will outline the concept of the dose-response relationship, and explain its application in this study of PA and behavioural engagement.

### 2.1.7 Dose-Response Relationship

'A dictionary of psychology' defines the dose-response relationship as: "The functional relationship between the dosage of a drug and the size or strength of a physiological or behavioural response to it" (Colman, 2015). In the case of most drugs, the relationship is a positive, linear relationship, with a larger response for a greater dosage of the drug – within a certain range (ibid.) (Figure 2A). However, sometimes a minimum 'threshold' dose is required before effects are noticeable (Schmidt-Kassow *et al.*, 2013) (Figure 2B), or very large or very small doses may have the opposite effect to a medium dose (Colman, 2015) (Figure 2C).

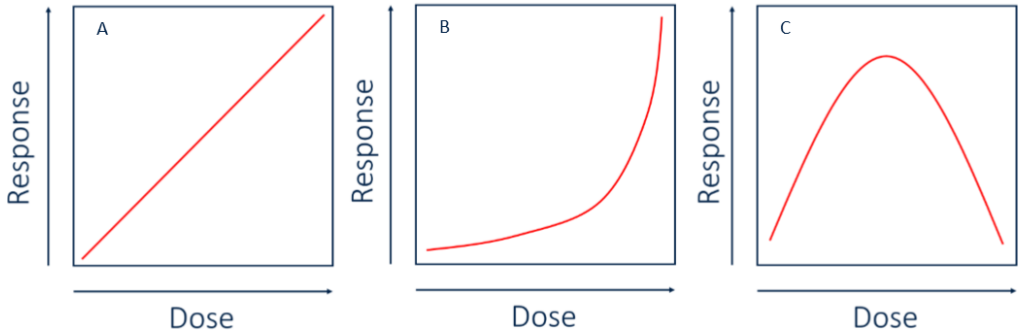



*Figure 2. Three different dose-response relationships. A. Linear, B. Threshold, C. Inverted-U.*

When considering the dose-response relationship between PA and learning behaviour in the context of this study, the dose (or the IV) is the level of intensity and complexity of the PE lesson. The response (or the DV) is the intended change in behavioural engagement. The task to be completed will influence the necessary dose of PA. For example, a task requiring gross motor skill is best

performed under high arousal (dose = vigorous PA), whereas tasks dependent on fine motor skill or complex cognitive performance are most successfully completed with low levels of arousal (dose = low-to-moderate intensity PA) (Schmidt and Wrisberg, 2004).

Moreover, there is the potential for conflicting responses stemming from the same type of exercise. One example of this conflict is vigorous PA (VPA). VPA leads to increased levels of cortisol (the "arousal/stress" hormone), which impairs hippocampal functions such as memory formation and retrieval (Weinberg *et al.*, 2014; Robbins, 2002). However, VPA also leads to arousal of the central nervous system (CNS) (Tomporowski, 2003), and increased expression of brain-derived neurotrophic factor (BDNF) (Knaepen *et al.*, 2010). Both have a positive effect on cognitive functioning due to their facilitating effect on neuronal functioning in the brain (Knaepen *et al.*, 2010; Schmidt and Wrisberg, 2004). Additionally, whereas memory retrieval is hindered by increased cortisol levels, memory consolidation is aided (Weinberg *et al.*, 2014; Hoetting *et al.*, 2016). How beneficial VPA is to task performance will therefore depend entirely on the specific nature of the task.

Thus, despite the consensus that PA is positively related to learning outcomes, the varied results found to date highlight that more research is needed into the exact parameters of this dose-response relationship. This study specifically investigates the relationship between PA intensity and complexity, and subsequent behavioural engagement.

### Curve Shifting

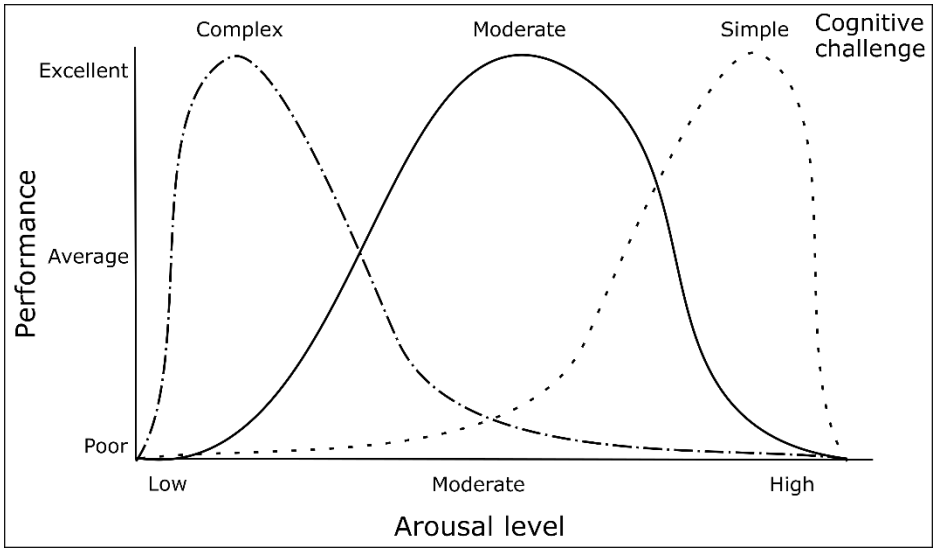



*Figure 3. Curve shifting as a result of task type (adapted from Schmidt & Wrisberg, 2004).*

An individual's characteristics, such as fitness level, may lead to curve-shifting in the dose-response relationship between PA and behavioural engagement (Schmidt and Wrisberg, 2004). Figure 3 demonstrates the principle of curve-shifting of the inverted-U curve of arousal and performance, based on the cognitive challenge presented by the task to be performed (ibid.). The amount of arousal (low, moderate, or high) which is deemed optimal for task performance depends on whether the task presents simple, moderate, or complex cognitive challenges. Curve-shifting also occurs between individuals who perform the same task. In that case, it is related to individual characteristics, such as fitness level or trait anxiety (ibid.). Section 2.3.7 explores the role of individual differences in the PA-cognition relationship in more detail.

This concludes the section on the concepts involved in this study. The following section presents existing theories of the relationship between PA and cognition, and the mechanisms by which PA affects cognition.

## 2.2 Theories of Physical Activity and Cognition

This section presents a range of theories on the mechanisms by which PA affects cognitive function and behavioural engagement. Lubans *et al.* (2016) propose a conceptual model which brings together many of these theories. They define cognitive function, well-being, and ill-being as the outcomes of exposure to PA, moderated by the PA content, through neurobiological, psychosocial, and behavioural changes (Figure 4). The present study incorporates the psychosocial and behavioural mechanisms: the

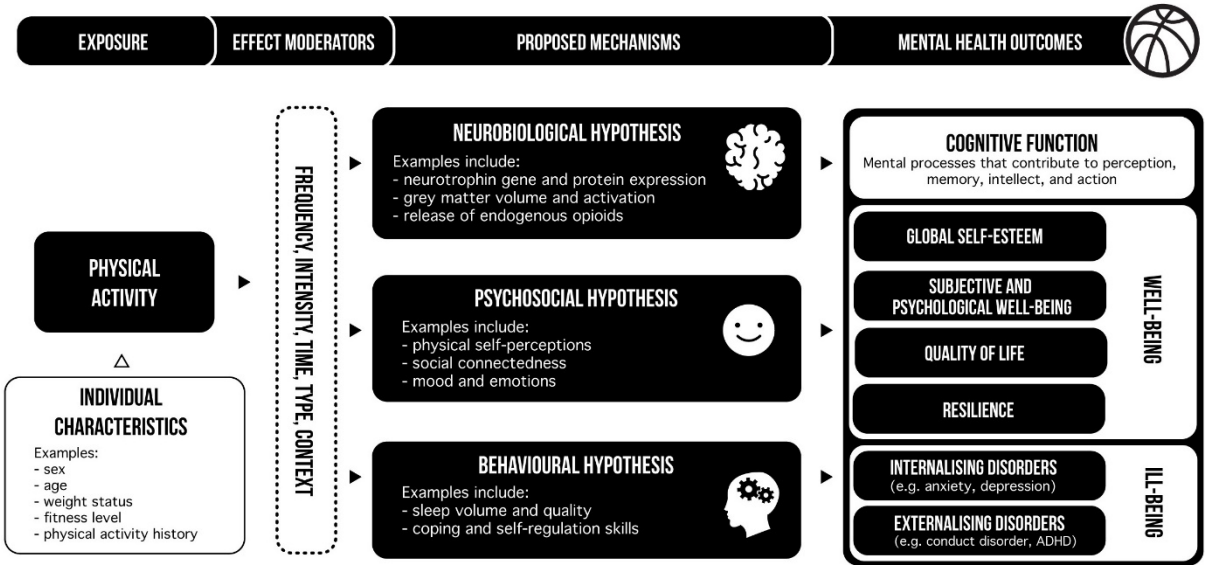



*Figure 4. Proposed conceptual model of PA and mental and cognitive health (Lubans et al., 2016).*

effect of PA on subsequent on-task behaviour was examined as well as the potential mediation effects of emotional state in the relationship between PA and on-task behaviour. Although no neurobiological measures were collected, due to this study taking place in a naturalistic setting, this section commences with a description of the neurobiological mechanism and arousal theory. This is because behavioural engagement is a behavioural manifestation of the neurological processes which can be affected by PA. Next, the psychosocial mechanisms are discussed, including the effects of PA on emotion states. This section concludes with an explanation of the behavioural mechanisms by which PA can affect cognitive functions, including Novelty Theory and the Play-Deprivation Theory.

### 2.2.1 Neurobiological Mechanisms

It has long been known that in mammals, including exercise as part of their lifestyle leads to increased neuronal growth and improvement to the neural systems involved in learning and memory (Chaddock *et al.*, 2011; Hillman, Erickson and Kramer, 2008). It has been established that sustained exercise leads to increased volume in several brain regions in both animals and humans (Brockett, LaMarca and Gould, 2015; Hillman *et al.*, 2008). Correlations have been found between a person's fitness level and the volume of their grey and white matter, as well as the density of blood vessels in the brain. These measures of brain volume and structure have in turn been linked to cognitive performance. For example, the relationship between physical fitness and relational memory performance in children has been established to be mediated by bilateral hippocampal volume (Chaddock *et al.*, 2011); and Erickson, Hillman and Kramer (2015), in a systematic review, describe how superior executive and memory function are related to greater brain connectively in physically fit samples versus lower-fit ones, both in childhood and older adulthood.

With regards to the acute effects of PA in rodents (rather than increased fitness), during and immediately after a bout of exercise, the blood flow in the brain is increased (Brockett *et al.*, 2015). This provides additional oxygen, glucose, and exercise-induced neuromodulators. Oxygen and glucose demands are increased in active neurons (ibid.), and thus in increased demand during learning activities. Neuromodulators increase the human brain's processing efficiency, including attentional allocation (Budde *et al.*, 2008). Therefore, the increased circulation of these substances as a result of PA facilitates learning.

Specific neurobiological systems affected by acute exercise are those involving norepinephrine (NE), epinephrine, and dopamine. NE and epinephrine lead to arousal of the neuronal structures in the brain. The role of NE in attentional allocation was discussed in section 2.1.5. Acute exercise induces an increase in the levels of NE in humans (Hoetting *et al.*, 2016; Nieuwenhuis *et al.*, 2005), and therefore affects alerting with an inverted-U relationship. Epinephrine expression in young adults is also boosted by acute PA (Winter *et al.*, 2007). Epinephrine increases noradrenergic release, leading

to greater neuron and brain excitability, and therefore improved learning capacity (ibid.) and alertness (Posner *et al.*, 2007). More details on arousal theory are provided in the next section.

Dopamine, which again is enhanced by PA (Winter *et al.*, 2007), has been linked to cognitive performance in a variety of manners. It is part of the intrinsic reward system and of emotional regulation (ibid.), and as described in section 2.1.6, a more positive emotion state may be beneficial to learning behaviour. Additionally, dopamine regulates the prefrontal cortical processes involved in working memory function. Finally, it modulates arousal and attention (ibid.); the dopaminergic system does not affect alerting or orienting (Davidson and Morrocco, 2000), but does affect attentional control (Posner *et al.*, 2007). Together, the exercise-induced changes in concentration of these three substances relate to alerting, attentional control, and emotional responses in ways that, at the right doses, can be facilitative to learning and attention.

### Arousal Theory

Arousal is related to both physiological and psychological states of being awake and alert (Byun *et al.*, 2014). The level of arousal is the extent to which neuronal tissue is activated, even if no neurological impulses transition as a direct result of the arousal. Arousal experienced by exercise study participants is thought to be one of the pathways by which acute PA affects cognition (ibid.). This is related to the increased circulation of norepinephrine and epinephrine as a result of PA participation. Exercise-induced arousal can stem from both physiological and psychological experiences during the exercise (Schmidt-Kassow *et al.*, 2014). Especially high-intensity exercise causes psycho-physiological stress, leading to an increase in arousal (ibid.). This arousal of nervous tissue during exercise can pre-activate the neuronal structures in the central nervous system (CNS), including the brain, leading to more efficient and effective processing in subsequent cognitive tasks (Budde *et al.*, 2008). This is observable in greater processing speed and better accuracy (ibid.).

Arousal theory is based on the overlap in the brain areas involved in motor tasks and in cognitive processing for learning and attention, so that exercise leads to stimulation of these overlapping

neuronal structures ahead of the learning episode. The main brain areas involved are discussed in the next sections of this chapter. The way in which pre-stimulation benefits subsequent cognitive processing is attributed to long-term potentiation, and a lowering of the threshold for action potentials, two processes upon which cognitive function depends. These processes are described in more detail in appendix E. According to arousal theory, exercise makes the brain more efficient at the neuronal level in the processes involved in learning.

Arousal has been established to affect attentional allocation with an inverted-U relationship (Schmidt-Kassow *et al.*, 2013). With increasing levels of arousal, perceptual narrowing takes place, a process described in the cue-utilisation theory (Schmidt and Wrisberg, 2004). In low arousal situations, attention is paid to all stimuli – the performer is easily distracted by irrelevant stimuli, and easily misses a relevant stimulus due to the overload of irrelevant stimuli (ibid.). In moderate arousal conditions, as perceptual narrowing takes place, the performer perceives only those stimuli relevant to performance and other stimuli are ignored, making the performer less distractible (ibid.). Finally, in high arousal circumstances, extreme perceptual narrowing occurs and the performer misses important stimuli which are required for optimal performance, leading to an observable decline in performance level (ibid.). In case of extreme arousal, hypervigilance may occur, better known as panic, and severely limited decision-making capacity, as well as a lack of smoothness in physical execution, is experienced (ibid.).

Another possible explanation for the negative effects of high arousal on cognition is the potential for dehydration and exhaustion directly after high-intensity activity (Hoetting et al., 2016). Dehydration can manifest itself as feelings of tiredness, lack of alertness and stimulus identification, slower reaction times, and poorer memory performance (Winter *et al.*, 2007). Optimal cognitive functioning occurs when arousal is high enough to avoid distraction by irrelevant stimuli, but focus remains wide enough to attend to all relevant stimuli (Schmidt and Wrisberg, 2004).

## Brain Areas Involved in Exercise and Cognitive Functioning

The neurobiological mechanism of pre-stimulation is reliant on the notion that the same areas of the brain are utilised in both exercise and the cognitive functions required during the learning task. PA affects many areas of cognition, and is evident in a range of areas in the brain (Hillman *et al.*, 2008). Additionally, exercise is not limited to motor action, and most forms of PA involve a cognitive demand. For example, cognitive demand can stem from the need for decision making and the application of tactics in sport (Schmidt and Wrisberg, 2004). Those forms of exercise which include a cognitive component, or highly complex movements, have been highlighted as a particularly effective source of neuronal pre-stimulation.

*Cerebral hemispheres.* The cerebral hemispheres are part of the forebrain (Figure 5), and further divided into four lobes (Romero, 2006) (Figure 6). Event Related Potentials (ERP, a type of brain imaging) have shown strong sensitivity to PA and fitness in the frontal lobe, infero-temporal lobe and parietal cortex (Hillman *et al.*, 2008). These areas, involved in stimulus processing, are important for learning and are indicative of better performance in tasks requiring attentional resource allocation (ibid.). Neuronal structures in the frontal lobe are related to both coordinative (motor) and cognitive processing (Budde *et al.*, 2008). Frontal lobe structures are particularly vital to processes involved

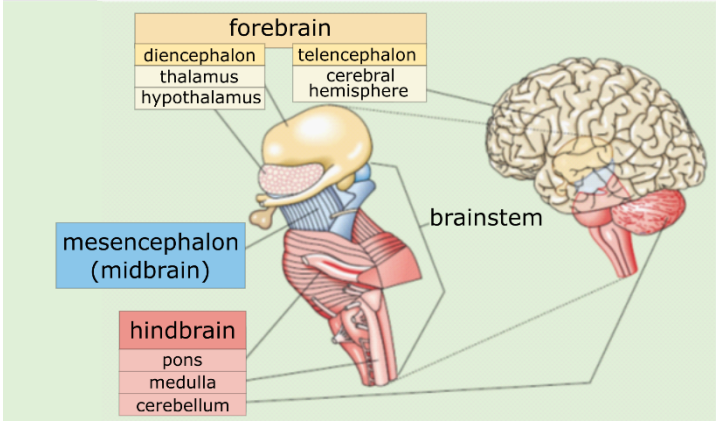



*Figure 5. The human brain, with divisions between forebrain, midbrain, and hindbrain indicated (Romero, 2006:7).*

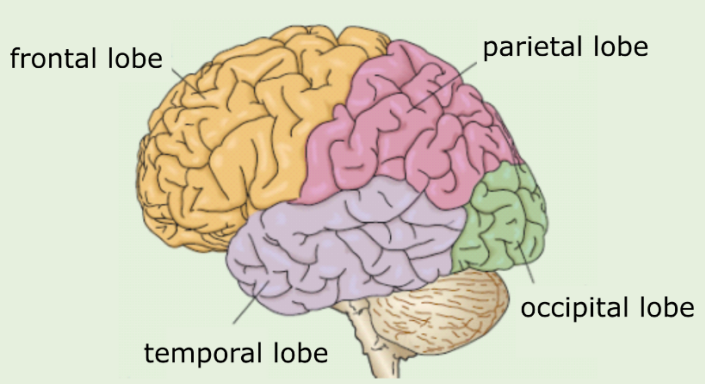



*Figure 6. The four lobes of the cerebral hemispheres (Romero, 2006:103).*

in classroom learning, as the lateral prefrontal cortex is involved in the application of knowledge in novel contexts (Posner *et al.*, 2007). Furthermore, the frontoparietal network is linked to language and numerical processing, with activity registered in this area of the brain during mathematics tasks as well as reading (Hillman *et al.*, 2008). The cerebral hemispheres, and more specifically the frontal lobe, are of particular interest in this study due to their involvement in both PA and EF.

*Cerebellum*. The cerebellum is located in the hindbrain (Figure 5). The cerebellum has long been known as a centre in the brain for motor functions (Steinlin, 2007). Additionally, in more recent years, several important cognitive functions have been related to the cerebellum, making it an especially interesting area for the link between PA and cognition. The main areas of cognition located in cerebellum are executive, visuospatial and verbal functions, and cerebellar defects have been linked to deficits in attention, processing speed, language, and visuospatial functions such as depth perception (ibid.). Increased cerebellar activity has been registered by functional Magnetic Resonance Imaging (fMRI) during tasks completed by novices as compared to experts, indicating that the cerebellum is particularly involved during new or unfamiliar tasks (ibid.). It follows that in a school-based environment, where students are continually introduced to new situations and tasks, the cerebellum is especially important for academic success. Like the frontal lobe, the cerebellum is a particularly interesting area of the brain for the link between PA and attentional allocation.

### 2.2.2 Psychosocial Mechanisms

Together with the behavioural hypothesis, the psychosocial hypothesis in the model by Lubans *et al.* (2016) accounts for the non-physiological effects of PA. Included in the psychosocial part of the model are subjective well-being, mood and emotions, self-perception, and anxiety and depression (ibid.). PA participation is put forward as a mechanism to increase self-esteem and positivity, and to reduce the experience of negative emotions, as PA provides opportunities for social interaction, skill improvement and mastery, and improvement in physical appearance (ibid.). They do however caution that only well-designed and -delivered PA will reap these rewards.

A systematic review of the psychosocial effects of PA by Eime *et al.* (2013) led to the conclusion that organised team sports, in particular, were beneficial to mental health and subjective well-being, due to their social nature. Another special mention in the psychosocial benefits literature is given to PA in natural environments. Spending time in the outdoors, specifically in natural environments, has been associated with positive mental health outcomes (Capaldi, Dopko and Zelenski, 2014), so it may be that PA in those environments is especially beneficial.

Of special interest in the context of acute PA are the psychosocial effects which occur instantly; those which are related to emotions rather than mood. Frequently mentioned benefits of PA participation are increased feelings of happiness, greater positive affect, feeling more energetic, and reduced anxiety (Eime *et al.*, 2013). Thus, it is possible that some of the effects of PA on cognition are related to an altered experience of emotions and stress in the classroom.

### 2.2.3 Behavioural Mechanisms

The behavioural mechanisms proposed in the model by Lubans *et al.* (2016) include the facilitation of behaviour more conducive to learning through good sleep patterns, improvements in self-regulation, and increased focus as a result of PA participation. As increased task-focus is related to academic achievement (section 2.1.512), and self-regulation skills are required to maintain attention (Pekrun *et al.*, 2002), PA-induced increases in self-regulation would benefit academic outcomes.

The maintenance of attention has long been a focus of the educational literature. Established theories on learning behaviour and attention specify the need for breaks to maintain focus (Berlyne, 1966; Smith and Hagan, 1980). More recent work has focused on PA breaks during the school day to maximise the benefits of these breaks for time-on-task and task-focus. The following sections set out two theories of learning behaviour; the first explains the benefits of breaks for learning behaviour and attention (novelty theory), and the second specifies the need for PA as well as social interactions during these breaks to maximise their benefits (play-deprivation theory). Together, these two theories

underpin the behavioural mechanisms by which PA breaks in the school day can contribute to improved learning and attention.

### Novelty Theory

Berlyne's novelty theory is based on the human tendency to get bored in unchanging environments, and the need for novel stimuli to maintain alertness (Posner *et al.*, 2007; Berlyne, 1966). The central nervous system never ceases to be active when a human is awake (Berlyne, 1966); levels of alertness fluctuate through the day, ranging from sleep-like states to highly vigilant ones (Posner *et al.*, 2007; Schmidt and Wrisberg, 2004). Throughout the day, the sense organs are constantly presented with stimuli (Berlyne, 1966). When naturally occurring stimuli are either too taxing or not demanding enough, humans seek their own stimuli through exploration (ibid.). This type of exploration is most likely to take place after spending time in a monotonous environment, or one lacking in stimulation (ibid.).

This is relevant in the classroom as insufficient stimulation or challenge, or indeed too much of it, can increase pupils' tendency to become bored or disengaged from the lesson, and lead to off-task behaviours. Another possible cause of novelty-seeking behaviour is task duration; when a task's duration is long enough for habituation to occur, perceived stimulation will decline and novelty will be sought to alleviate boredom (Jarrett *et al.*, 1998; Berlyne, 1966). When this aspect of novelty theory is applied to the classroom, it can be expected that when behavioural engagement declines as a function of time, breaks can serve to prevent or remedy habituation to tasks (Pellegrini and Davis, 1993). The level of excitement sought by individuals depends on their personality, culture, psychophysiological state and experience (Berlyne, 1966). The psychophysiological state refers to the optimal level of arousal, as previously discussed in relation to arousal theory and PA. Based on these theories, if breaks have a PA content they can be utilised during the school day to provide psychophysiological arousal and maintain students' engagement in their lessons. This will be further discussed in the next section, which will set out the play-deprivation theory.

### Play-Deprivation Theory

Novelty seeking in humans is motivated also by their innate desire for learning and increasing knowledge (Berlyne, 1966). The play-deprivation theory assumes that children's natural form of learning is through physically active and social play (Pellegrini and Holmes, 2006; Smith and Hagan, 1980). Activities which limit PA and social interactions, and interfere with the unstructured nature which is inherent in play (Pellegrini and Holmes, 2006) are described as 'play deprivation' (Pellegrini and Holmes, 2006; Smith and Hagan, 1980). Classroom lessons are such periods of 'play deprivation'. These periods of deprivation lead to a 'rebound', consisting of increased PA and social interaction (Smith and Hagan, 1980). Prolonged deprivation negatively affects concentration (Pellegrini and Smith, 1993; Pellegrini, Huberty and Jones, 1995; Smith and Hagan, 1980), thus explaining children's need for breaks during the school day.

In conclusion, both the novelty theory and the play-deprivation theory imply the need for breaks as part of the structure of children's school-based learning. As well identifying the need for variation within tasks to maintain alertness and avoid a lack of attentional allocation, breaks which contain PA and social interactions are suggested as a necessity for maintaining behavioural engagement in the classroom. Such breaks can take the form of a PE lesson, or consist of free play such as takes place at morning break and lunch time in most primary schools (Heemskerk, 2015). The exact nature of the dose-response relationship between the PA content of the break, and subsequent attentional allocation is still unknown. An overview of recent research in this area is presented in section 2.3.3. This study aims to contribute to the clarification of the dose-response relationship between PA and subsequent behavioural engagement, and possible mediating effects of emotional states.

## 2.3 Research Evidence on Physical Activity and Cognition

Overall, evidence of the benefits of PA for cognitive function is positive, although results are varied and sometimes contradictory (De Greeff *et al.*, 2018; Watson *et al.*, 2017). Variation stems from the

PA parameters as well as the cognitive outcomes selected, creating a complex landscape of causal and correlational links. This section aims to give an outline of the relationships between PA and cognitive functions and academic achievement. It will do so with particular attention to chronic PA versus acute PA, and on behavioural engagement, mood and emotion states as outcome variables. Another focus is on research conducted in PE lessons. Where possible, the emphasis will be on results from research conducted with child samples and in school environments, though this literature remains sparse compared to (young) adult samples and laboratory settings. This study contributes to this specific type literature in the field of PA and cognition.

### 2.3.1 Chronic Physical Activity and Cognition

Research has linked chronic PA and the resultant improvements in fitness to several aspects of cognitive function. Many cross-sectional studies have been conducted into the relationship between (aerobic) fitness levels and various cognitive functions and/or academic achievement. Additionally, intervention studies involving participation in long-term PA have started to uncover a cause-effect relationship between chronic PA, fitness, and cognition. Overall, a positive relationship between chronic PA and cognitive function has been uncovered.

In their analysis of correlational studies (k=20), Fedewa and Ahn (2011) established that children with greater physical fitness performed better in cognitive tests (d=0.32). Consistent with that meta-analysis, in a cross-sectional study of 259 seven to ten-year-olds, Castelli *et al.* (2007) found that overall fitness and aerobic fitness had positive correlations with mathematics achievement scores (d=1.01, p<0.001), reading achievement scores (d=0.82, p<0.001), and overall academic achievement (d=0.95, p<0.001). And Chomitz *et al.* (2009), in a sample of 1478 eight to thirteen-year-olds, found a similar relationship between fitness and mathematics scores (d=1.12, p<0.0001), with a slightly weaker relationship for English scores (d=1.01, p<0.0004).

However, in a review of 39 studies, Watson *et al.* (2017) included measures of standardised test scores as well as teacher ratings of achievement, and found that only teacher assessments increased

with increased participation in PA (d=1.03, 95%CI[-0.22,1.84]). Likewise, Coe *et al.* (2006) found no increased performance in 214 10 to eleven-year-olds on standardised test scores for mathematics, English, Science, or World Studies in relation to PE lesson participation. Nevertheless, those who achieved the recommended vigorous activity levels set out by the 'Healthy People 2010' guidelines, significantly outperformed those who did not at both timepoints (first semester: $\chi^2$=10.1, p<0.006, d=0.45; second semester: $\chi^2$=6.05, p<0.049, d=0.35) (ibid.). This may be an indication that increased participation in PA per se does not correlate with academic performance, but that certain other criteria about the PA need to be satisfied, e.g. the duration of VPA.

A wide range of literature has investigated the relationships between fitness and isolated cognitive skills, rather than scholastic achievement. The majority of recent literature is on memory and executive functions (EF) such as inhibition or updating. In a meta-analysis of 19 studies focused on long-term participation in exercise, Roig *et al.* (2013) established a small but significant relationship with short-term memory (d=0.15, 95%CI [0.02, 0.27], p=0.02), and overall memory (d=0.15, 95%CI [0.04, 0.25], p=0.006), but not with specific measures of long-term memory. Pontifex *et al.* (2011) used ERPs to investigate attentional allocation in 48 lower- and higher-fit children (M age = 10.0). Medium to large effect sizes (d's ranging 0.6-1.04) in the amplitudes (indicating allocation of attentional resource) and latencies (reaction time to a stimulus) of all investigated measures indicated that attentional allocation in lower-fit children is less efficient, and more response conflict occurs in this group (ibid.).

Buck *et al.* (2008) found specific correlations between aerobic fitness, as measured with the PACER test, and EF performance (selective attention, response inhibition, and interference control) in the Stroop task in healthy 7-12-year-olds. Controlling for age and IQ, those with higher levels of aerobic fitness outperformed lesser fit children in all conditions (word: d=0.82, p=0.001, colour: d=0.56, p=0.02, colour-word: d=0.65, p<0.01), making fewer errors in naming the word/ink colour. No interference effect was found, indicating that performance requiring greater inhibitory control was not disproportionately affected by fitness (ibid.).

In contrast to the findings from Buck *et al.* (2008), EF has been implicated as a mediator between physical fitness and academic achievement by Van der Niet *et al.* (2014). Working with 243 children aged 7-12, they used structural equation modelling (SEM) to investigate the relationships between physical fitness, EF, and academic achievement. In isolation, both EF and academic achievement were significantly related to physical fitness (d=0.95 and d=0.70 respectively). However, when EF and academic achievement were both entered into the same model with physical fitness, the direct relationship between fitness and academic achievement was no longer significant, whereas the relationship mediated by EF was (d=0.90) (Van der Niet *et al.*, 2014). This would suggest that observed positive correlations between physical fitness and cognitive outcomes are due to improved EF in higher fit children.

Finally, measures of fitness not only correlate with academic achievement and cognitive functioning, they also influence how individuals respond to acute exercise (Chang *et al.*, 2014). Individuals' fitness (low, medium, high) was found to have a large effect on the response to a bout of acute PA in relation to Stroop task performance (d=1.19, $p<0.001$), with medium fitness optimal. These results imply that even when investigating effects of acute exercise, individuals' fitness level cannot be ignored. With that in mind, section 2.3.2 will now give an overview of the current literature on acute PA and cognition.

### 2.3.2 Acute Physical Activity and Cognition

Like chronic PA, various parameters influence the effects of acute PA on cognitive outcomes, and the effects further fluctuate depending on the cognitive measure used. This section will give an outline of the research investigating the effects of duration, intensity, and type of exercise on cognitive task performance.

Chang *et al.* (2015) identified twenty minutes as a more effective duration of PA than ten or 45 minutes, in relation to cognitive tasks. PA opportunities in schools tend to vary between ten-minute 'active breaks', to around one hour for lunch time and PE lessons. However, it must be noted that

although PE lessons in UK primary schools tend to be 50-60 minutes long, 40-45 minutes is typically spent in the sports hall or on the playground/field. The same applies to lunch break, of which a large proportion is spent queueing for food and eating, leaving around 30 minutes of PA time. Additionally, children have different PA profiles from adults, with more intermittent bursts of activity (Schaefer *et al.*, 2014). These bursts, lasting on average three to four seconds each, can be accumulated to achieve the total activity duration (Schaefer *et al.*, 2014).

With regards to PE lessons, the child-specific nature of exercising in bursts, together with the structure of PE lessons (time for instructions, taking turns) will affect the accumulated duration of PA. In a 45-minute lesson, children will not be active for 45 minutes; Coe *et al.* (2006) found that as little as nineteen minutes of a 55-minute PE lesson was spent at moderate-to-vigorous PA (MVPA). Consequently, lessons with more time spent waiting to take turns or listening to instructions will accumulate less activity time and children's perceived exertion and tiredness will be more similar to a lower-intensity lesson, even if the time spent being active was at a vigorous intensity.

The intensity of acute PA is said to have an inverted-U relationship with cognitive functioning whereby the greatest benefits are attributed to moderate intensity PA (MPA) (Tomporowski, 2003). Although there is increasing evidence indicating benefits of chronic VPA (e.g. Coe *et al.*, 2006), the reported effects of acute VPA are inconsistent, and vary depending on the cognitive task. Covassin *et al.* (2007) found negative effects of high-intensity PA on immediate ($d=0.56$, $p<0.001$) and delayed ($d=0.45$, $p<0.001$) verbal memory, but Salis (2013) reported increased vocabulary comprehension after VPA, but not MPA ($F(2,66)=4.82$, $p=0.03$, $d=0.77$). For reaction time the results are also mixed. Alves *et al.* (2014) found improvements in reaction time on the Stroop colour-word condition after bouts of high-intensity PA ($d=-0.53$, $p<0.001$). Yet, Tsai *et al.* (2014) reported reduced reaction times on a visuo-spatial task after 30 minutes of acute MPA ($d=0.89$, $p<0.001$). Thus, VPA may not necessarily be better than MPA.

Exercise type, as well as exercise duration and intensity, and cognitive task type, influences the effect of PA on cognition. In a recent meta-analysis, De Greeff *et al.* (2018) found no significant effect

for acute cognitively engaging PA, but a small to moderate effect for acute aerobic PA (g=0.43, 95% CI [0.09, 0.77]). Similarly, Donnelly *et al.* (2016) found in their systematic review that studies of acute PA, and especially those in school settings, varied widely in the type of PA used and the results yielded. For example, Chen *et al.* (2014), investigating 30 minutes of moderate intensity running, found a positive effect on response time in three EF tasks: inhibition (d=0.81, p<0.001), working memory (d=0.91, p<0.001) and shifting (d=0.91, p<0.001). Comparing playful aerobic exercise to team games, Schmidt *et al.* (2015b) found that shifting performance was improved only in the games group (d=0.22, p=0.012).

Cognitive task performance has been found to be affected by acute PA, with results established to depend on the duration (Chang *et al.*, 2015), type of exercise (Donnelly *et al.*, 2016; De Greeff *et al.*, 2018), and intensity (Tomporowski, 2003; Covassin *et al.*, 2007). In the context of this study, duration cannot be varied as PE lessons are of a set duration. Instead, the content of the PE lessons will be manipulated to change the intensity and type of exercise. Moreover, the results of studies of PA and cognition depend on the chosen outcome measures. In this study, the outcome variable is behavioural engagement, an important antecedent of academic achievement (Watson *et al.*, 2019). Literature reporting on specific findings about the relationship between PA and behavioural engagement is presented in the next section.

### 2.3.3 Physical Activity and Behavioural Engagement

In the 1990s and early 2000s, research focused on the effects of break times in primary schools on on-task behaviour, based on the belief that children need breaks in order to maintain focus in the classroom. In an intra-personal experimental study using direct observation, Jarrett *et al.* (1998) found that 4th graders were on-task more (F(1,42)=10.02, p=0.003 d=0.97), and displayed less non-task-related action (F(1,42)=31.36, p<0.001 d=1.72) on days when they had a break compared to non-break days. Likewise, in a cross-sectional study of 11,624 eight- and nine-year-olds, Barros, Silver and Stein (2009) found that teacher ratings of learning behaviour were significantly more positive in schools

where children had breaks, compared to those with minimal or no breaks (d=0.18, p<0.001). However, neither study qualified the type of behaviour children engaged in during their break, so these results cannot be attributed to PA participation.

One of the studies that did quantify break time behaviour was the work of Pellegrini and Davis (1993), which used an intra-personal design to investigate 3rd grade pupils' concentration and fidgeting in class during long (3h) and normal (2.5h) classroom instruction. They observed PA and social interactions at break time, and related these to the duration of play deprivation, as well as subsequent classroom behaviour. Their findings confirmed the hypothesis that longer deprivation led to more vigorous active play at break time. Additionally, they observed that vigorous exercise play during break was correlated with increased subsequent fidgeting (d=0.30), whereas more sedate, social behaviour was correlated with better concentration (d=0.28) and less fidgeting (d=-0.35). In a similar study design, Pellegrini *et al.* (1995) again found that longer deprivation led to a decrease in attentiveness, and that attentiveness was improved by break time. However, no significant correlation was identified between the rate of social interactions during break time, or the intensity of children's PA behaviour, and subsequent classroom attentiveness. Thus, although the various study results indicate that breaks are related to increased attentiveness during subsequent lessons, no dose-response relationship with PA or social content was identified.

In more recent years, the research focus has turned to PA breaks, whereby all pupils engage in a certain amount of teacher-directed PA. These breaks can take place either in the classroom, or in the school hall/outside. PA breaks have gained in popularity across primary schools in the United Kingdom and across the world in the past decade. Examples of initiatives for short duration daily PA breaks implemented in UK primary schools are 'wake up, shake up', a five to ten minute PA programme suitable for the classroom (Wake Up Shake Up Productions, 2014), and 'the daily mile', which gets children to go outside and run or jog for 15 minutes every day (The Daily Mile Foundation, 2019). Mahar (2019) reviewed a range of classroom PA initiatives from the US, Canada, and The Netherlands, and found relatively consistent positive effects on on-task behaviour (d=0.31-2.2).

Positive results were reported by Carlson *et al.* (2015), after 1322 children (mean age 8.8 years, SD=1.5) participated in activity breaks. Fewer instances of lack of effort in class were reported by teachers (β=-0.17, p=0.042), and duration of MVPA during school the school day negatively associated with teacher ratings of students being off-task or inattentive (β=-0.17, p=0.042) (ibid.). Another positive result of PA breaks on time-on-task was reported by Howie, Beets and Pate (2014), who found that 96 children in grades 4 and 5 were more on task after ten minutes of PA break compared to a sedentary break (d=0.45, p=0.004). Whitt-Glover, Ham and Yancey (2011) executed ten-minute PA breaks set to music, and despite finding that light PA (LPA) duration increased more than MPA duration did (51% and 16% respectively), they also observed an 11% increase in time-on-task in the classroom after the PA breaks (p=0.038). Finally, Mahar *et al.* (2006) implemented 'Energisers', a daily intervention consisting of ten minutes of classroom-based PA for 3rd and 4th grade pupils. Not only did they find that on-task behaviour improved (d=0.6, p<0.017), but that it improved dramatically more in those children who were the least on-task before Energisers took place (d=2.2, p<0.001). This implies that there are individual differences in how PA breaks affect children's learning behaviour, consistent with the curve-shifting principle discussed in section 2.1.7.

Although no negative effects of PA breaks on subsequent learning behaviour have been reported in the literature, not all studies found positive effects. Wilson *et al.* (2016) found no significant effect of a ten-minute active lesson break outside the classroom on on-task behaviour in 58 boys (mean age 11.2 years, SD=0.6). Unlike the previous studies reported, the participants in the study by Wilson *et al.* (2016) did not perform the PA in the classroom. Schmidt, Benzing and Kamer (2016) found no effect of PA during classroom breaks on subsequent focussed attention in 98 5th grade students. The difference in their results compared to the studies discussed before may lie in the use of the d2-R task to assess focussed attention immediately after the intervention, rather than direct observations. It is therefore possible that the behavioural benefits of the PA break went unnoticed as the assessments focused on cognitive rather than behavioural measures. Moreover, the assessments took place over a much shorter period of time, as the d2-R took only four minutes to complete, compared to direct

observation over a period of up to 30 minutes in other studies. More details of the d2 and d2-R task can be found in appendix C.

Altogether, the evidence of the effects of PA breaks on subsequent learning behaviour is positive, and short duration bouts of PA of around 10 minutes are an easy way to improve children's time-on-task, in particular in those who struggle the most with behavioural engagement. Effect sizes found in studies of PA breaks are similar to, or greater than, those reported in studies of break time, suggesting that PA during the breaks has no negative effect, and possibly a positive effect, compared to less active breaks. The content of the PA breaks varies between intervention programmes; they vary in intensity, location, and some, like Whitt-Glover *et al.* (2011), use exercise set to music. Another type of exercise that has been rapidly gaining ground in recent years is cognitively engaging exercise, and an increasing number of PA interventions are examining the specific benefits of this type of exercise. The following section will discuss cognitively engaging exercise breaks and their relation to behaviour engagement and measures of attention and EF.

### 2.3.4 Cognitively Engaging Physical Activity Breaks

As the qualitative content of PA continues to be much investigated, research on cognitively engaging PA has increasingly gained a place in the literature, as evidenced by it being mentioned as a category in its own right in meta-analyses and systematic reviews (De Greeff *et al.*, 2018; Watson *et al.*, 2017). In a meta-analysis of 31 studies on acute PA and cognitive functions in pre-adolescent children, De Greeff *et al.* (2018) concluded that there was an overall small to moderate positive effect of acute aerobic PA (6 studies) on selective attention (Hedges' g=0.43, p=0.013), but no significant effect of cognitively engaging PA (5 studies). The lack of an overall effect of cognitively engaging PA may have been due to the diverse findings reported in the literature. A large variety exists in the time period over which the effects were measured (immediately after and up to 50 minutes later), as well as in the outcome measures included in studies, though these have tended to be psychometric tests of EFs. Details of three frequently used psychometric tests can be found in appendix C.

Best (2012) carried out an experiment with 33 six- to ten-year-old children taking part in PA, cognitively engaging PA, a sedentary cognitive challenge, or a resting control condition with no PA or cognitive challenge. They reported a positive main effect of PA on interference control in a flanker task immediately after the intervention, but no main effect for cognitive engagement, or interaction effect of PA and cognitive engagement. In an experiment with 104 slightly younger children, aged 6-8, Jäger *et al.* (2014) assessed measures of updating, shifting, and inhibition before, immediately after, and 40 minutes after a 20-minute intervention. The intervention was delivered to children in groups of four and involved cognitively engaging PA or a resting control where they listened to a story. Gains were found in inhibition measures on the modified flanker task for the PA group both immediately after, and 40 minutes after the intervention, with a slight reduction in the effect after 40 minutes ($d=0.49$, $p=0.02$ and $d=0.45$, $p=0.03$).

Working with 217 slightly older children (mean age 11y4m, SD 6.4m) Jäger *et al.* (2015) carried out a 20-minute intervention in more naturalistic settings. Participants attended their intervention activity (PA, cognitively engaging PA, sedentary cognitive challenge, or resting control) in a group setting in their school. There was no main or interaction effect of PA or cognitive challenge on shifting, updating, or inhibition for the whole sample. This suggests that PA-induced improvements in cognitive function may be harder to achieve or detect in a naturalistic environment.

Another explanation is that individual differences are important to consider. Jäger *et al.* (2015) did find a main effect of PA on updating for higher fit children, as higher fit children in the cognitively challenging PA and aerobic PA groups outperformed those in the sedentary cognitive challenge group ($d=0.79$ and $d=69$, respectively) and those in the resting control group ($d=0.59$ and $d=0.50$, respectively). There was also a main effect of PA and of cognitive challenge on updating for higher achieving children. Higher achieving children benefited from aerobic PA ($d=0.89$, $p<0.01$), cognitively engaging PA ($d=1.12$, $p<0.01$), and sedentary cognitive challenges ($d=0.57$, $p=0.02$). Lower fit and lower achieving children did not respond positively or negatively to any of the intervention conditions.

Budde *et al.* (2008) found that d2 task performance improved in 115 thirteen- to sixteen-year olds after participation in a ten-minute bout of PA. However, there was a significant interaction effect of condition and time, with those in the coordinative exercise group outperforming their peers in the heart rate matched control group performing aerobic exercises on all measures (items completed ($d=0.37$, $p<0.01$), ratio correct : errors ($d=0.44$, $p<0.01$), % errors committed ($d=-0.39$,$p<0.01$)). Their study was conducted in an elite-performance school and will have included only higher-achieving participants, which may have influenced their results.

The studies discussed in this section examined the effects of short-duration cognitively engaging PA on cognitive functions. These types of PA breaks have been popular with schools and teachers as they do not take up much classroom time and are easy to integrate into the school day. However, in primary schools in the United Kingdom, there is also a timetabled PA opportunity in the form of PE lessons. These longer periods of PA have been the focus of some researchers, and manipulations of the content of PE lessons have been investigated in relation to cognitive function. The following section will discuss several studies which have compared the effects of cognitively engaging PE lessons with intensity-matched aerobic lessons.

### 2.3.5 Physical Education Interventions

Few studies have been conducted into the effects of manipulated PE lesson content on measures of cognition. The studies reported in this section have examined the effects of PA and cognitively engaging PA during PE lessons on a variety of psychometric measures. PE lessons are of a longer duration than the PA breaks reported in the previous sections, and duration of PA has been indicated as a possible factor that influences the dose-response relationship with cognitive function (Chang *et al.*, 2015). It is therefore important to consider these studies separately from the literature on PA breaks.

Schmidt *et al.* (2015b) implemented a 6-week PE programme in which classes were assigned to a low intensity aerobic PA group (control), a high intensity aerobic PA group, or a high intensity,

cognitively engaging PA group. Participants were 181 children aged ten to twelve. Both of the high intensity PA groups significantly improved their aerobic fitness, but only the cognitively engaging group recorded improvements to measures of EF; shifting performance on the flanker task improved compared to the control group (d=0.22, p=0.012) and compared to the aerobic exercise group (d=0.17, p=0.039). There was no effect of the intervention on updating and inhibition. This intervention investigated chronic effects of specific PE lesson content. Next, studies of the acute effects are discussed.

In a sample of 52 students aged 11-12, Pesce *et al.* (2009) examined verbal recall after 40-minute PE lessons consisting of either aerobic PA (aerobic circuit) or cognitively challenging PA (team games), or without exercise beforehand. They reported that delayed recall was improved after both types of PE lesson (team games: d=0.86, p=0.003, aerobic circuit: d=1.10, p<0.001), but immediate recall was improved only after the team games (d=0.68, p=0.019 (first five of the 20 items) and d=1.25, p<0.001 (last five of the 20 items)). Memory performance was consistently worst when not preceded by exercise (ibid.).

Gallotta *et al.* (2015) implemented a PE intervention consisting of 50-minute PE lessons involving either aerobic, or cognitively engaging aerobic activities of matched intensity, and a control group which took part in a 50-minute classroom lesson instead of PE. Participants (116 eight- to eleven-year-old children) completed the d2 task before, immediately after, and 50 minutes after the PE lesson. There was one group-by-time interaction effect, which revealed that the percentage change in the number of correct items minus the errors committed in post-test and 50-minute follow-up was lower in the cognitively engaging PA group than in either of the other groups (ibid.). The aerobic PA group outperformed the cognitively engaging PA group on attentional control after the PE lesson. This indicates that cognitively engaging PE lessons may not be as beneficial as aerobic PE lessons for attentional control. Considering that the classroom-based group performed significantly worse than either PA group at baseline (ibid.), the relative improvement of the classroom-based control group might not be an accurate reflection of the effects of PA versus classroom time on attentional control.

Evidence from studies on full-duration PE lessons is sparse, and results vary by outcome measures. Whereas verbal memory and shifting task performance seems to be improved after cognitively engaging exercise, measures of attentional control have been indicated to benefit from aerobic rather than cognitively engaging PE lessons. This study contributes to this body of literature by systematically comparing exercise intensity and cognitive demand.

### 2.3.6 Physical Activity, Emotion States, and Behavioural Engagement

In this section, the research evidence in support of the psychosocial mechanisms is examined. First, evidence is presented of the relationship between PA and emotion states. Next, studies of emotion states and their influence on learning, cognitive function, and attentional allocation are discussed. Finally, the these are brought together to create a mediated link between PA and behavioural engagement.

#### Physical Activity and Emotion States

PA is generally found to positively affect mood and emotions in both children and adults. Long and Van Stavel (1995) reported positive effects of PA on anxiety in adult samples (d=0.45 for within-person designs, d=0.36 for between-person designs). In a child sample (n=70, M age=10.4) Parfitt and Eston (2005) found effect sizes for anxiety of greater magnitude than those reported in adults (d=-1.09).

The effect of PA on mood and emotion states has been determined to depend on several PA parameters such as intensity and duration (Chan *et al.*, 2018). In their systematic review they conclude that MPA is most effective at improving mood, yet both Crews, Lochbaum and Landers (2004) and Strauss *et al.* (2001) found that those children participating in more VPA had greater gains in measures of self-esteem. Strauss *et al.* (2001) identified significantly higher self-esteem ratings specifically in children who participated in more VPA (d=0.58), and Crews *et al.* (2004) reported improvements for the experimental group (>60% HR intensity, d=-0.39), but no effect for the control group (50% HR intensity). Crews *et al.* (2004) also found that only in the vigorously active group measures of

depression declined (d=-0.24). Rather than intensity, Chekroud *et al.* (2018) found that the modality of PA affected the magnitude of the effect on mental health outcomes. Although all forms of exercise were associated with a lower mental health burden, especially beneficial were popular team sports, cycling, and aerobic activities (ibid.).

Overall, participation in chronic PA was found to be effective at reducing negative mood (Chan *et al.*, 2018). Regular participation at a frequency of three to five times per week, as well as durations of around 45 minutes have been highlighted to be the most effective at improving mood and reducing poor mental health symptoms (Chekroud *et al.*, 2018), although some suggest that ten to fifteen minutes is sufficient (Chan *et al.*, 2018). Results reported from chronic PA interventions with child samples include teacher-reports of increased positivity in pupils after daily PA (Macdonald *et al.*, 2014), and reduced measures of depression with increased weekly step-count ($F(2,64)=13.86$, $p<0.01$, $d=0.93$) (Parfitt and Eston, 2005).

Acute PA has also been reported to have positive effects on emotions, mood and affect. A PA intervention in depressive adults including indoor and outdoor PA and an indoor control group found that both PA groups reported significantly higher positive affect (indoor: d=0.64, outdoor: d=1.44) (Frühauf *et al.*, 2016). In a study of 108 females in a community fitness centre setting, Gauvin, Rejeski and Norris (1996) collected self-reports on emotion states within 20 minutes before and after vigorous exercise classes for six weeks. Positive engagement, tranquillity, and positive affect all increased after PA compared to before PA ($d=0.91$, $p<0.001$, $d=1.01$, $p<0.001$, and $d=1.34$, $p<0.001$ respectively), and negative affect significantly decreased ($d=-0.47$, $p<0.01$) (ibid.).

Acute bouts of PA have been suggested to induce positive mood for as long as 30 minutes after completion of the bout (Chan *et al.*, 2018). In line with this assertion, Dunton *et al.* (2014) reported that higher levels of MVPA in the 30 minutes before the completion of a self-report were associated with higher ratings of positive affect and feeling energetic, and lower ratings of negative affect in 119 children aged nine to thirteen.

## Emotion States and Attentional Allocation

Recent studies by Pöysä *et al.* (2017), Martin *et al.* (2015), and Vasalampi *et al.* (2016) have identified within-person fluctuations in (behavioural) engagement on a lesson-to-lesson basis, even within the same day. Suggested reasons for this include pupils' subject enjoyment, teacher factors, and instructional format. Another possible cause, which will be examined in this study, is pupils' emotional state (Scrimin and Mason, 2015).

Emotion states can either increase task-focus or induce task-irrelevant thinking that can be detrimental to task-focus (Pekrun *et al.*, 2002). In a study of 230 university students, positive task-intrinsic emotions have been found to correlate with reduced task-irrelevant thinking ($r=-0.38$, $p<0.001$), whereas in the case of negative emotions both task-intrinsic or task-extrinsic emotions were related to increased task-irrelevant thinking, with the strength of the correlation depending on the specific emotion ($r=0.41$-$0.72$, $p<0.001$) (Titz (2001), in: Pekrun *et al.*, 2002).

Meinhardt and Pekrun (2003) recorded ERP's to test the assumption that negative emotions function as a drain on attentional resources which would otherwise be allocated to the task. They found that this is the case not just for negative emotions, but also for positive ones. University students had ERPs recorded whilst they completed an oddball task (a dual processing task) and had positive and negative emotions induced by seeing emotionally laden pictures or by reading scripts about emotional events. In both scenarios, P3 amplitude (a reliable indicator of processing resources in the brain) was reduced when participants were induced to have heightened emotion states (picture-induced: $d=0.81$, $p<0.001$, script-induced: $d=0.97$, $p<0.001$), indicating a reduced allocation of resources to the oddball task. There was no difference between positive and negative emotion states (picture-induced: $d=0.25$, $p>0.05$, script-induced: $d=0.01$, $p>0.05$).

In both studies, participants were university students and emotions were induced by academic tasks in a learning environment (Pekrun *et al.*, 2002), or artificially induced with visuals (Meinhardt and Pekrun, 2003). My study investigates emotions which are induced or modified by PA with or

without cognitive demands, and how these emotion states mediate the effect of PA on behaviour in the classroom.

One study was identified in which the mediating role of affect was investigated in the relationship between focussed attention and school-based PA breaks with cognitive challenge. Working with 92 fifth-grade pupils, Schmidt *et al.* (2016) found that changes in focussed attention (measured with the d2-R) after the cognitively challenging condition of the ten-minute intervention were only significant when mediation effects of positive affect were included (β=-0.03, p=0.02). No direct or mediated effects were found for PA (ibid.).

Together, these results show that both PA and cognitive demands can affect emotion states, and that emotion states influence cognitive functions such as attentional allocation. Evidence of mediating effects of emotion states in the relationship between PA and cognitive function is scarce. My study aims to contribute to the literature by investigating emotion states in relation to both classroom behaviour and to six systematically manipulated combinations of PA intensity and complexity for the duration of a full PE lesson.

### 2.3.7 The Role of Individual Characteristics in the Physical Activity – Cognition Relationship

As briefly discussed in section 2.1.7, individuals respond differently to PA based on their personal characteristics, with the most clearly established ones being age (e.g. Buck *et al.*, 2008) and fitness level (e.g. Donnelly *et al.*, 2016; Chang *et al.*, 2014). This section will provide more details on a range of individual characteristics which have been identified as potential moderators in the relationship between PA and measures of cognition.

#### Aerobic Fitness Level and Trait Physical Activity Participation

The relationship between aerobic fitness and cognitive function is well-established. For example, Chaddock *et al.* (2011) found an association between low levels of PA and poor aerobic fitness, and lower academic achievement and cognitive ability. ERP studies show positive association between

aerobic fitness and processes occurring between stimulus and response execution (ibid.). Lower aerobic fitness is related to reduced amplitude in P3, indicating reduced neural inhibition and thus reduced facilitation of attentional processing in the brain (ibid.). Lower aerobic fitness also relates to slower latency of the P3 signal, indicating reduced evaluation and classification speed (ibid.). Thus, lower aerobic fitness relates to delayed processing and classification and reduced capacity to inhibit superfluous neural operations (ibid.), evident in a breakdown in attention in those with poorer aerobic fitness (Hillman, Castelli and Buck, 2005). Furthermore, Pontifex *et al.* (2011) found that those with higher aerobic fitness were better able to maintain high achievement across changing task demands. In adults, higher-fit individuals registered greater activation of the prefrontal cortex and parietal cortex during an interference task than lower-fit individuals, and task performance indicated greater inhibition control in the higher-fit individuals (Buck *et al.*, 2008). A similar result was found in preadolescent children, with those with better fitness scores outperforming those with lower levels of physical fitness in a Stroop task (ibid.).

Lab-based studies generally involve one participant at a time, exercising at a customised intensity to achieve the target HR/$VO_2$max percentage (Tomporowski, 2003; Schmidt-Kassow *et al.*, 2014). This is not a realistic situation for a school PE lesson, as classes consist of individuals with a range of fitness levels, and individual differences in perceived exertion are unavoidable. This may have knock-on effects in the individual response to the PA dose: individuals who are higher-fit will likely feel less exhausted by the same PE lesson than lower-fit individuals, and the peak of their arousal curve will occur at a different lesson intensity (Schmidt and Wrisberg, 2004). With regard to trait anxiety, those with high trait anxiety are predisposed to perceive external pressure as stressful, and they are likely to respond differently to an externally paced exercise condition than those with low trait anxiety, who naturally seek high levels of arousal and perceive such a situation as a challenge (Berlyne, 1966; Schmidt and Wrisberg, 2004).

Fitness is a result of chronic PA participation, so it is not surprising that chronic PA participation also affects the response to PA. Coe *et al.* (2006) found that those children who achieved the 'Healthy

People 2010' guidelines for vigorous activity outperformed those who did not achieve the guidelines in standardised tests of academic achievement. Watson *et al.* (2019) reported that those children who did not participate in organised sports were less attentive (individual sports B=-1.00 and team sports B=-0.88) and more impulsive (individual sports OR=0.35) than those who did.

### Age, Gender, and Body Mass Index

Another individual characteristic to consider is age, as pre-adolescent brains are still developing (Buck *et al.*, 2008). Whereas fMRI results in adults show that the effects of fitness are greater in those tasks with a greater EF demand (Hillman *et al.*, 2008), this relation was not observed in preadolescent children (Buck *et al.*, 2008). One possible reason for this may be that, in preadolescent children, the frontal lobe area of the brain is not yet fully developed, and their ability to inhibit is lower than in adults. This implies that study results from adult samples cannot be generalised to children, and that age, as well as fitness and trait anxiety, needs to be considered in the investigation of acute PA and attentional allocation.

Although the majority of studies report controlling for gender, gender-based differences in response to PA are rarely reported (e.g. Watson *et al.*, 2017). Yet, as boys are generally found to spend less time on-task in lessons (e.g. Bugler, McGeown and St Clair-Thompson, 2015; Skinner *et al.*, 2008; Martin, 2007), and those with the lowest levels of on-task behaviour have been identified as benefiting the most from PA interventions, there is scope for boys to book greater gains in on-task behaviour after PA than girls. One study which did find gender-based effects is the work of Ma, Mare and Gurd (2014), who found that in boys, but not girls, passive off-task and motor off-task behaviours were reduced after PA breaks.

In a study of children above the 85$^{th}$ body mass index (BMI) percentile for their age and gender, Davis *et al.* (2007) reported positive effects of a 15-week long daily after-school dose of PA (40-minutes per session) on performance in a planning task ($d=-0.26$, $p<0.05$). However, as there was no healthy-weight control group, it is not possible to conclude if these results differ on the basis of BMI status.

Results from Grieco, Jowers and Bartholomew (2009) suggest that, in inactive classroom lessons, overweight children experience a greater decline in time spent on-task as lesson duration increases than those of a healthy weight. They also found that in active classroom lessons, this decline does not occur in children of any BMI status, concluding that the effect of PA is magnified for overweight children (ibid.).

### Trait On-Task Behaviour and Attainment

Mahar *et al.* (2006) found that those who were least on-task during baseline observations improved their on-task behaviour the most after taking part in PA ($d=2.02$, $p<0.001$). They defined the 'least on-task' students as those who spent less than 50% of the pre-break lesson on-task. It intuitively makes sense that those with the lowest baseline measures have the greatest scope for improvement, but this still is an important point for investigation. As described in section 2.1.5, besides being prohibitive to the attainment of the students who are off-task (e.g. Watson *et al.*, 2019), the disruption caused by off-task behaviour affects the learning of all students in the class (Mahar *et al.*, 2006; Godwin and Fisher, 2011). Teachers are therefore concerned about the behaviour of those students who have low levels of behavioural engagement, and PA may provide a practical tool to assist them in preventing disruptive behaviour.

Jäger *et al.* (2015), found that only those with attainment scores in the top half of the sample experienced positive effects of PA, on updating. However, shifting and inhibition were unaffected by attainment level (ibid.). In light of this finding, the results from Budde *et al.* (2008) should be interpreted with caution as those were achieved in an elite performance school, and this may have influenced the effect sizes reported.

## 2.4 Physical Education and Physical Activity in Primary Schools in the United Kingdom

As is the case in many places in the western world, PA opportunities in UK primary schools are declining (Morgan and Hansen, 2008a). National and international comparison tables with a focus on literacy and mathematics have added pressure to school administrators and teachers to produce high grades in standardised tests (Tsangaridou, 2017). Subjects such as PE, which have a lower socio-political status related to a lack of accountability, have been marginalised as a result (Andrew *et al.*, 2016; Tsangaridou, 2017). In the UK specifically, PE was dropped from the curriculum between 1998 and 2000 to make more time for academic subjects (Harris *et al.*, 2012). Upon reintroduction of PE to the UK primary school curriculum, the government established a target of 2h of high-quality PE and school sports for all primary school children. In 2006, the DfES reported that 80% of pupils were achieving the 2h target. However, a survey of 459 primary school teachers conducted by Harris *et al.* (2012) revealed that more than 75% of pupils were not receiving 2h a week. This discrepancy may be due to timetabled PE not always taking place; teachers report often scrapping PE in favour of academic subjects (Harris *et al.*, 2012). Reasons provided by primary school teachers include a lack of confidence and insufficient content knowledge (ibid.).

PE has suffered from low status in initial teacher training (ITT) programmes, with 25% of primary ITT reported to spend less than 10h on PE instruction (ibid.), and some programmes spending as little as 4h (Freak and Miller, 2017). This has led to many primary school teachers feeling ill-prepared to teach PE. Morgan and Hansen (2008a) found that the likelihood of teaching PE is related to the teacher's attitude to PE and their belief in its importance ($d=0.49$, $p<0.01$), and this was found to be significantly higher in male teachers than in females ($t(180)=2.25$, $p=0.03$, $d=0.16$). Primary school teachers in the UK are predominantly female, as represented in the sample of Harris *et al.* (2012), which was 82% female. Even for those teachers who believe PE is important, low confidence in their own ability to teach and assess progress in PE, and concerns about pupils' safety remain often-cited

barriers to teaching (Tsangaridou, 2017; Morgan and Hansen, 2008a; Harris *et al.*, 2012; Freak and Miller, 2017).

These barriers to teaching PE can be removed through high quality ITT ($d=0.75$, $p<0.001$), especially for games ($d=1.62$, $p<0.001$) (Freak and Miller, 2017). Teacher reports of confidence and attitude are generally higher in studies conducted with samples of specialist PE teachers (positive attitude reported by >90% of the sample) (Rachele *et al.*, 2016), and those on ITT programmes which dedicate a greater amount of time to PE; 84.5% of pre-service teachers reported feeling prepared for teaching PE after a minimum of 39h compulsory PE instruction (Freak and Miller, 2017). Another method of removing barriers of subject knowledge is high quality professional development (PD) for in-service teachers: Iserbyt, Ward and Li (2017) found that as a result of teacher participation in additional training, the quality of instruction increased, and errors in pupils' skilled performance reduced ($F(1,48)=48.18$, $p<0.001$, $d=2.21$), whilst correct trials increased ($F(1,48)=59.89$, $p<0.001$, $d=2.00$). The results from my study can be used to inform ITT and PD courses on the wide range of benefits of PE lesson content to pupil development. Its contribution to teacher knowledge can assist in creating a more positive attitude towards PE in primary school teachers.

In the next section, gaps in the current literature are outlined, and the way in which this study will address these gaps is highlighted. Following this, a conceptual model of the dose-response relationship of PA, emotion states, and behavioural engagement is introduced, based on existing models of the relationship between PA and cognition, and the hypotheses of this study.

## 2.5 Gaps in the Current Literature

The literature has the most pertinent gaps in the dearth of studies of pre-adolescent children and of those conducted in naturalistic settings. The vast majority of PA studies are conducted with young adults or with aging participants. Few studies have been conducted with adolescents, and fewer still

with younger children. As the brain structures in pre-adolescents are still developing – and in particular the frontal lobe, a structure heavily involved in inhibition and behavioural engagement, and central in the current study – results from older samples cannot be generalised, and it is important that research is conducted with this age group. As well as a dearth of age-appropriate studies, there is a need for more studies into the acute effects of PA on behavioural engagement. A large body of research has investigated the relationship between fitness and achievement, between fitness and attentional allocation, and between acute PA and performance on a range of cognitive tasks. However, to the best of my knowledge, studies into behavioural manifestations of the effects of PA in schools remain sparse, as research on acute PA in schools is typically related to academic achievement or cognitive tests.

Situating the research in schools rather than in a laboratory will automatically address the earlier issue of age-appropriateness, but brings other challenges. Most pertinent is the level at which analysis takes place. Neuroscientific research investigates learning starting at the micro-level of molecules, and working up to neurons, neuronal networks, and finally behaviour. Educational research on the other hand, uses behaviour at its lowest level, and places this in various levels of context, from lessons and classrooms up to schools and countries. This means that one cannot be generalised to the other, and both Bruer (1997), and more recently Bowers (2016) argue against integrating educational and neuroscientific research. But, if the field of education is to make practical use of the wealth of knowledge available in neuroscience, research which is based on principles and results from laboratories needs to be embedded into the wider context in which school-based learning take place, and be tested for its applicability. Results achieved in laboratories, without application in classrooms, are of no use to teachers.

This study was situated in classrooms, and normal classroom lessons took place before and after the PE intervention lesson. The strength of this approach lies in the increased ecological validity of the study, and making it applicable to real-life situations. On the other hand, it exposes the intervention

to many contextual influences which cannot be controlled, and leads to a trade-off of decreased external validity for increased ecological validity (Coolican, 2009).

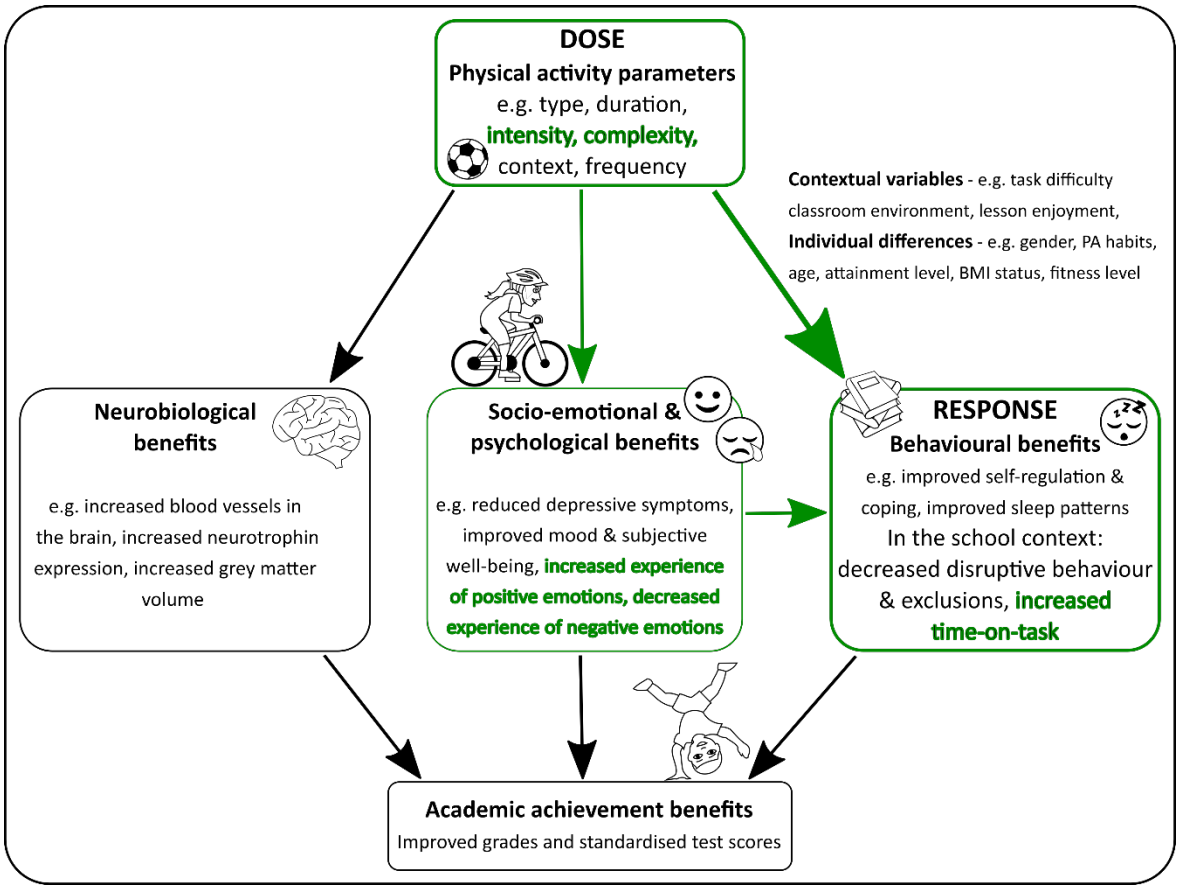



*Figure 7. Proposed conceptual model of the pathways by which PA benefits academic achievement.*

Note: sections highlighted in green were investigated in my empirical study.

Drawing on the existing literature, the conceptual model presented in Figure 7 is proposed. This model theorises that PA affects academic achievement via the neurobiological, psychosocial, and behavioural mechanism proposed by (Lubans *et al.*, 2016). An addition in this model is the specific differentiation between behavioural benefits that affect academic performance through improved general well-being (e.g. sleeping patterns) and those particular to the school/academic context (e.g. classroom behaviour). A further proposed addition is that the behavioural benefits are mediated by the socio-emotional and psychological benefits, as indicated by the horizontal green arrow. The parts of the model which were tested in this study are highlighted in green: the dose-response effect of PA

intensity and complexity on subsequent on-task behaviour in the classroom, and the mediation effect of emotion states in the relationship between PA and on-task behaviour.

## 2.6 Research Questions and Hypotheses

To address the gaps in current research, this study will focus on the complexity of exercise and the aerobic intensity of exercise. The research questions for this study are:

1. Does exercise complexity affect subsequent levels of behavioural engagement during learning tasks?
2. Does the aerobic intensity of exercise affect subsequent levels of behavioural engagement during learning tasks?
3. Is there an interaction effect of aerobic intensity and complexity of PA on subsequent levels of behavioural engagement during learning tasks?
4. Is there a specific sub-group of the sample for whom exercise content more strongly affects subsequent behavioural engagement during learning tasks?
5. Do emotions mediate the relationship between PA and behavioural engagement during learning tasks?

Based on the existing literature, it is hypothesised that:

1. Exercise with high complexity/cognitive demand will increase subsequent behavioural engagement more than exercise with low complexity/cognitive demand (Figure 8A, section 2.3.3)
2. The intensity of aerobic exercise will have an inverted-U relationship with subsequent behavioural engagement (Figure 8B, section 2.3.4)

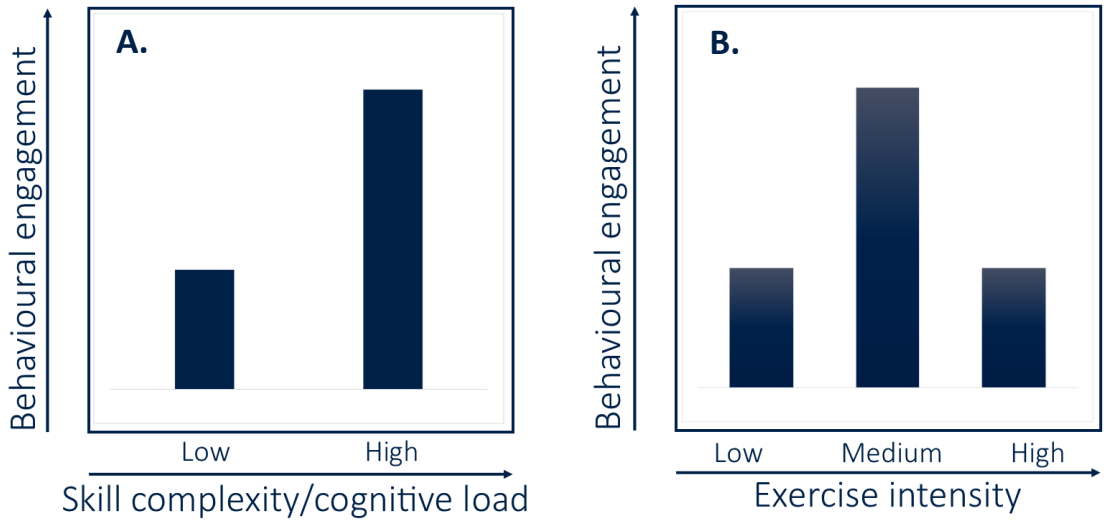



*Figure 8. The hypothesised effects of skill complexity (A) and exercise intensity (B) on subsequent behavioural engagement.*

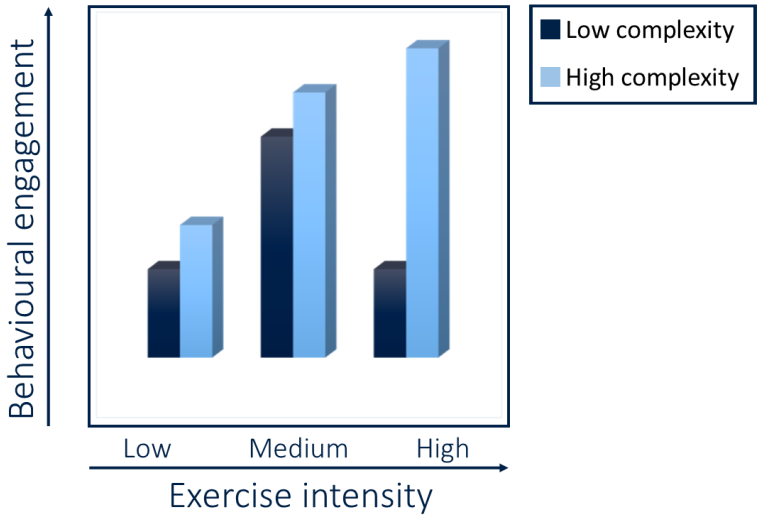



*Figure 9. The hypothesised interaction effect of skill complexity and exercise intensity.*

3. There is an interaction effect of complexity/cognitive load and intensity, whereby high complexity will change the inverted-U relationship of intensity so that vigorous intensity activity will be followed by a stronger increase in behavioural engagement than moderate intensity activity (Figure 9, section 2.3.4).
4. Individual differences in gender, trait behavioural engagement, PA participation, and academic achievement will influence the intervention effects. Greater improvement of behavioural engagement after PE is expected for those participants who are male, who are the least on-

task in pre-PE lessons, who take part in less PA, and who are higher academic achievers (sections 2.1.7 and 2.3.7).

5. Emotions mediate the relationship between PA and behavioural engagement (Figure 7, section 2.3.6).

# 3 Methodology

This chapter aims to set out the design of the current study, and its position within the field of PA and cognition research. First, the design of the study, and the methods and instruments chosen and developed for the data collection in this study are described. A rationale is provided for their selection, adaptation and/or development. Then, the sampling strategy is specified, and the sample is described. Next, the matters of validity, reliability, and research ethics appropriate to this study are discussed. Finally, the analysis strategies adopted for answering the research questions posed in the previous chapter are explained.

The current literature on PA and cognition comprises a huge range of aspects of PA, and a correspondingly large range of methods is used to investigate these. Laboratory-based studies tend to focus on precise measures of exercise intensity or fitness, based on heart rate or maximum oxygen capacity ($VO_2$max) (Westfall *et al.*, 2017; Hoetting *et al.*, 2016). The majority of studies on fitness are cross-sectional, relating participants' fitness level (which is defined differently in different studies) to their cognitive performance in a wide range of tests; from standardised tests of EF (Westfall *et al.*, 2017), to school achievement (Van der Niet *et al.*, 2014), to brain scan results (Pontifex *et al.*, 2011). Unsurprisingly, the range of results is equally large, although overwhelmingly positive.

This generally positive outcome also holds true for studies of acute exercise. The predominant methodologies employed in studies of acute PA are randomised controlled trials and repeated measures experiments. Again, laboratory-based studies form the majority of the literature, investigating the effects of exercise intensity (e.g. Alves *et al.*, 2014; Salis, 2013) and duration (e.g. Chang *et al.*, 2015) on biological markers such as blood and salivary hormone levels known to affect cognitive task performance (e.g. Winter *et al.*, 2007; Weinberg *et al.*, 2014), as well as vocabulary recall and outcomes on standardised test of EFs.

The considerable majority of research into PA and cognition is conducted with samples of adults or older adults (e.g. Hoetting *et al.*, 2016; Chang *et al.*, 2015; Alves *et al.*, 2014). It is recognised that these results cannot be generalised to younger populations; their developing brains respond differently to cognitive demand, so it cannot be assumed their cognitive response to exercise will be the same as a declining brain, or a healthy, fully developed brain (Guiney and Machado, 2013). Studies with child and adolescent samples include experimental as well as cross-sectional designs, and have been conducted in laboratories as well as naturalistic settings, mostly schools. IVs range from participation in active school lessons and extra-curricular PA to fitness levels and, in a handful of studies, modified demands of PE lessons. Outcome measures mostly involve academic achievement markers (teacher assessed or standardised tests) (e.g. Haapala *et al.*, 2016; Chomitz *et al.*, 2009) and tests of EFs (e.g. Berchicci *et al.*, 2015; Aadland *et al.*, 2017).

As indicated in section 2.3, although the literature is expanding, there still is a dearth of research with child samples and in naturalistic settings. Even those studies conducted in schools often use standardised tests of cognitive functions as their outcome measures, rather than naturally occurring learning behaviour. Lubans *et al.* (2016) only found two studies reporting on behavioural effects of PA interventions participation, and De Greeff *et al.* (2018) included six in their meta-analysis. More studies have been conducted, but these mostly examine the effects on behaviour of classroom-based PA breaks (Howie *et al.*, 2014; Szabo-Reed *et al.*, 2017) or active lessons (Bartholomew *et al.*, 2018; Grieco *et al.*, 2016). These PA breaks are generally of short duration and vary in intensity. There are multiple studies investigating the effect of PA on academic achievement, using teacher's assessments or standardised test scores. However, what these studies fail to address is the mechanisms by which the PA may affect cognitive outcomes. Is it because the way children respond to cognitive demand is altered by PA? Is it because exercise leads children to have different subjective experiences in the classroom, with different emotional states? Or does PA affect their behaviour, assisting them to be more efficient, less distracted learners?

Based on the model by Lubans *et al.* (2016) of the effects of PA on mental well-being and cognitive function in children and young people (Figure 4), a new conceptual model is proposed in this study, which suggests the three mechanisms work in parallel with each other, but also interact with one-another (Figure 7). This study contributes an investigation of the effects of PA intensity and complexity on learning behaviour. The main analysis considers on-task behaviour as a result of exercise intensity and complexity, and additional modelling will investigate if the relationship between exercise and on-task behaviour is mediated by self-reported emotional state (Figure 7, green elements).

## 3.1 Research Design

The research questions addressed in this study are concerned with causality: the effects of PA on behavioural engagement. In order to accurately assess the effect of various PA conditions, and claim causality, an experimental design was chosen (Bloom, 2012). The gold standard for experimental designs is considered to be the randomised control trial (Bloom, 2012; Cook and Wong, 2012). However, in the social sciences it is often difficult to achieve context-free research conditions (Sulkunen, 2012; Cook and Wong, 2012), as the phenomena researched are by definition embedded within a context. Although this clashes with the view of 'pure' science of the positivist stance, the close relationship with the context lends relevance to the study, and the value of research ultimately lies within its interpretation and application (Sulkunen, 2012).

The present study was conducted from the perspective of a class teacher and PE teacher. Thus, the emphasis is placed on behavioural engagement. Teachers are keen to see children engage in lessons, as it is well-documented that there is a link between engagement in lessons and academic achievement (Mahar, 2019). Pellegrini and Davis (1993) found that on-task behaviour affected academic achievement with an effect size of d=0.52. However, staying on-task in a classroom is challenging in itself; classroom lessons are subject to a host of distractors, and, unlike psychometric tests in controlled environments, these are not only present within the task content. Distractions are

presented by the other individuals in the class, possible visitors coming into the room, the weather outside the window, or any event that might happen outside such as an ambulance driving by with the sirens and blue lights on. To give a fair representation of the actual goings-on during lessons, this research was situated directly in the context of daily school life.

Instead of a (between-person) randomised control trial, a (within-person) longitudinal design was chosen. Longitudinal studies assess processes over time, and aim to identify the cause(s) of change or stability in phenomena (Elliot, Holland and Thomson, 2012). This is ideally suited for this study, as the stability as well as the prevalence of behavioural engagement over the duration of the lesson is of interest. Additionally, an intrapersonal design was chosen, with participants providing their own baseline-control of behavioural engagement before each PE lesson. Several recent studies have suggested intrapersonal research as a next step for assessing the benefits of PA due to the large range in individual differences in baseline measures (Hoetting *et al.*, 2016; Knaepen *et al.*, 2010). Heemskerk (2015) found large individual differences in the relationship between PA and on-task behaviour, suggesting that an intrapersonal approach is appropriate for this study. Godwin and Fisher (2011) also suggest that this approach is favourable as, by design, it controls for the effects of lesson/subject interest. More details on the inclusion of subjective ratings of task enjoyment as control variables can be found in sections 3.3.4 and 3.10.4. Section 2.3.7 gives more details of the effects of individual characteristics on the PA-cognition dose-response relationship.

## 3.2 Intervention Protocol

Despite the consensus that PA is positively related to learning outcomes, the varied results to date highlight that more research is needed into the exact parameters of this dose-response relationship. Cognitive task performance is affected by PA, with results established to depend on the duration (Chang *et al.*, 2015), type of exercise (Budde *et al.*, 2008; Schmidt *et al.*, 2015b), and intensity (Tomporowski, 2003; Covassin *et al.*, 2007). Duration cannot be varied in the context of this study, as

PE lessons are of a set duration. Therefore, in this intervention the content of six PE lessons was manipulated to change the intensity and complexity of the exercise.

The full intervention protocol can be found in appendix F. The six lessons which make up the intervention place varying demands on participants' skill complexity and cardio-vascular effort (Figure 10). In the left column are the low complexity lessons, with the lowest intensity at the bottom and the highest intensity at the top. Flexibility testing and stretching, health-related fitness, and sprinting games were low complexity lessons. Skill complexity in these lessons was kept low by using simple and routine movements and actions such as running (Best, 2012; Jäger *et al.*, 2015; Gallotta *et al.*, 2015). The flexibility test (the sit-and-reach test) required participants to sit on the floor in front of a bench and reach both hands as far forwards as possible, and similarly the hamstring stretch was carried out by reaching for one's toes in a seated position. In health-related fitness, participants completed a timed sprint, a standing long jump, and performed as many sit-ups and press-ups as they could in one minute, or to failure. All of these activities were made up of simple and isolated movements, most of which children routinely involve in their free play. In the sprint races and relays, children were encouraged to run as continuously as possible in a straight line which was marked out with cones. Running is a routine action for most children, and to keep the movement simple no focus was placed on developing running technique in this lesson. The relays were also performed in a

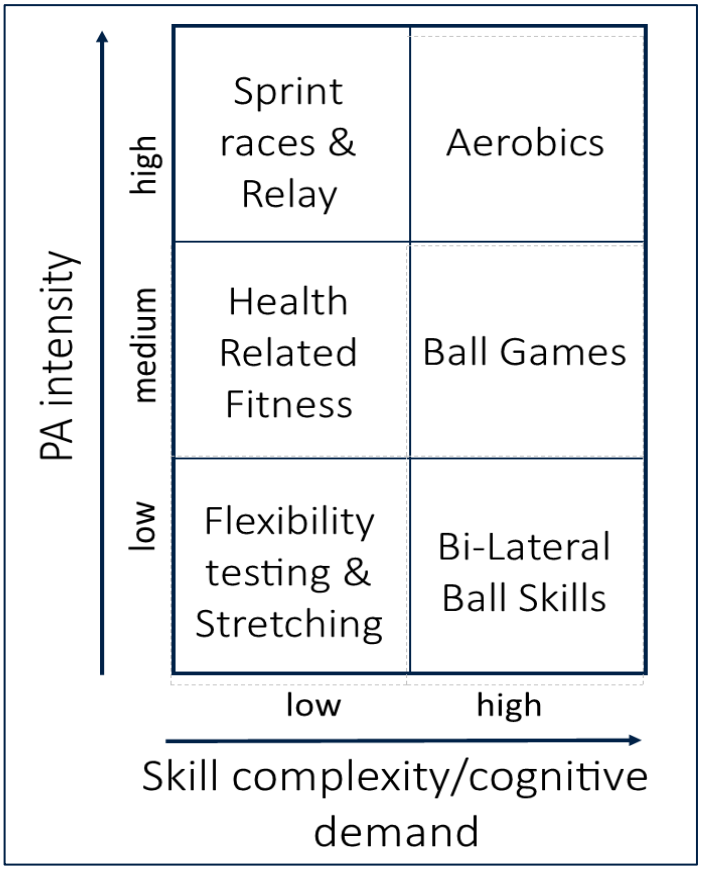



*Figure 10. Content of the intervention lessons*

straight line marked with cones, and to reduce the coordination required for the change-overs, partners were tagged instead of passing a baton.

In the right column of figure 10 are the high complexity lessons, again with the lowest intensity at the bottom and the highest intensity at the top. Bi-lateral ball skills, games, and aerobics were high complexity lessons. Skill complexity in these lessons was high as these activities required bi-lateral movements (Rusch and Irrgang, no date), quick reactions (Best, 2012), object manipulation (Rusch and Irrgang, no date), rule-changing (Schmidt *et al.*, 2015b), and/or timing (Gallotta *et al.*, 2015). The bi-later ball skills lesson centred around juggling and involved skill practices with complicated combinations of catching, throwing, and kicking. Ball games required skilled actions whilst on the move, and often with short reaction times in changing environments, whilst aerobics involved the coordination of all four limbs at the same time as well as timing to music and with other people. In both the low and medium intensity lesson, the children who were competent in the skills performed were challenged further by needing to use their non-dominant hand/foot for equipment handling.

The cardio-vascular demand was varied by the speed at which movements were performed (sprinting and jumping vs. jogging/stationary), and the continuity of movement (team games and taking turns vs. continuous movement for the whole class). In the low intensity lessons the participants were mostly seated or performed stationary/walking pace drills. In the medium intensity lessons the activities themselves were more intense, but intermittent as they needed to take turns or work in a team. In the high intensity lessons the actions were performed at high speed, and in the aerobics lesson also involved a high degree of jumping. Moreover, in these lessons the movement was made a continuous as was feasible considering the duration of the lesson, the intensity of the movements, and the age of the participants. Section 3.3.1 gives more details of the specific activities involved in each lesson.

Each of the six weekly 24h cycles of data collection started with participants being fitted with accelerometers at the end of a school day. The following day, the researchers arrived at school to set up the tablets for questionnaires during lunch time. At the end of lunch, once all children had returned

to the classroom and the class teacher started the lesson, the researchers started observations of six participants each, twelve in total. At five, 12.5, and 20 minutes after the start of observations, participants were prompted to complete a questionnaire. Observations finished after 25 minutes, and children then got ready to go to PE. Where possible, the children had gotten changed before lunch to make the transitions quicker, but this was not feasible in all schools. The PE lesson lasted between 40 and 50 minutes, and children completed a questionnaire at the end of the PE lesson, before entering the classroom. Once they completed their questionnaire, they returned straight to the classroom without getting changed. In the classroom, the class teacher taught another lesson, during which the same twelve participants were observed by the researchers. At five, 12.5, and 20 minutes after the start of observations, the participants were again prompted to complete a questionnaire. After 25 minutes, the researchers stopped observing, and collected the tablets and accelerometers from the participants.

## 3.3 Variables

This study investigates causality between PA and behavioural engagement. To accurately achieve this, the cause factor needs to be measurable and adjustable (Sulkunen, 2012). In this study, the cause factor is the content of the PE lesson, which was varied in six predetermined ways, affecting the skill complexity and/or the intensity of the activity (see section 3.3.1). In order to attribute the change in learning behaviour to the PE lesson content, other variables need to be eliminated as possible causes of the outcome, either empirically or by the use of control variables in statistical analyses (ibid.). As the current study was conducted in a naturalistic setting, empirical elimination of all other variables was not feasible, and instead background variables were collected alongside the baseline measures. More details of the control variables can be found in section 3.3.4. Finally, the effect must be measurable, which is achieved through the operationalisation of behavioural engagement, and by having a systematic observation protocol, described in section 3.3.2 and appendix C.

### 3.3.1 Independent Variable: Physical Activity

The PA content of the PE lesson was structurally manipulated in six ways, three intensity levels and two complexity levels, as shown in Figure 10. The following sections describe in detail how intensity and complexity were operationalised and manipulated in this study, and which PA variables were created on the basis of the measures collected during the intervention.

#### Exercise Intensity

The intended PA intensity of each lesson was included as a variable in the dataset (0 = 'low', 1 = 'medium', 2 = 'high'). Additionally, the PA intensity was objectively measured with the use of accelerometery devices (GENEActiv and Empatica E4). GENEActiv devices are widely used in PA research (Duncan *et al.*, 2016; Esliger *et al.*, 2011; Poh, Swenson and Picard, 2010) and have been validated for measuring PA in children, with cut-off points established for sedentary, low, moderate, and vigorous intensity activity (Phillips *et al.*, 2013; Schaefer *et al.*, 2014; Duncan *et al.*, 2016). More details on the devices and their validation process can be found in appendix B.

Objective, high-frequency measurement of PA is important with child samples as children have different PA profiles from adults, with more intermittent bursts of activity (Schaefer *et al.*, 2014). These bursts, lasting on average 3-4s each, are not easily detected by observation, but are recorded by the GENEActiv/E4 device which samples at a frequency of up to 100Hz (GENEActiv) or 32Hz (E4). Over the measurement period, the bursts are accumulated to achieve the total activity duration (Schaefer *et al.*, 2014).This is important, as children will not be active for the full duration of their PE lesson; Coe *et al.* (2006) found that as little as nineteen minutes of a 55-minute lesson was spent at MVPA. This child-specific aspect of the duration of PA is vital to accurately establish the intensity level of PE lessons, as lessons with more time spent waiting to take turns or listening to instructions will accumulate less activity time, and affect the body and mind differently to continuously active lessons. For example, the expression of hormones such as cortisol (Jäger *et al.*, 2014; Schmidt-Kassow *et al.*,

2014) and brain-derived factors (Hwang *et al.*, 2016; Schmidt-Kassow *et al.*, 2014) are affected by exercise intensity, as are subjective measures such as perceived exertion and subsequent tiredness. Accelerometery measures from PE lessons were used to check for intervention fidelity, see sections 3.10.4 and 4.2 for more details.

The devices were worn for 24h for each intervention day (up to 6 days) to provide an approximation of the free-living activity levels of the participants. Measures of an individual's chronic PA participation and level of fitness affect how they respond to acute exercise (Chang *et al.*, 2014). Fitness (low, medium, high) was found to have a large effect on the response to a bout of acute PA in relation to Stroop task performance ($d=1.19$, $p<0.001$), with medium fitness being optimal (ibid.). These results imply that even when investigating effects of acute exercise, individuals' fitness level cannot be ignored and needs to be controlled for. In this study no fitness testing could be conducted, due to time and resource constraints. Instead, recorded minutes of free-living MVPA per 24h were used to assess activity levels, with participants allocated to 'achieving >60min per day' or 'not achieving >60min per day', reflecting the current government guidelines for children's activity levels (Chief Medical Officers, 2011). The dichotomous allocation based on government guidelines was made for each day of recorded activity, and participants who achieved the guideline on at least 75% of their recorded days were termed 'active children'. The time spent in their PE lesson was excluded from the 24h free-living MVPA minutes as activity levels during these periods were dictated by the intervention rather than 'free-living'.

Two additional chronic PA categorisations were given to each participant based on self-reported participation in organised extra-curricular and out-of-school PA (referred to as 'organised sport'). More details on these additional variables can be found in sections 3.3.4 and 3.4.2.

### Skill Complexity

Skill complexity was classified by the complexity of the required movements. A variable was included in the dataset for the intended complexity of the lesson (0 = 'low', 1 = 'high'). To allocate

exercises and activities to either the low or high complexity categories, previous research was used as guidance. Budde *et al.* (2008) adapted exercises from the Muenchner Fitnesstest (MFT) for their intervention of coordinative exercise. In the MFT, the tests for coordination include balance, and alternate between left and right hand (Rusch and Irrgang, no date). The balance element was not adopted by Budde *et al.* (2008), but they modified tests to include greater bilateral demand for additional complexity. Schmidt *et al.* (2015b) varied the cognitive demand of exercises by varying rules, adapting the pace at which changes were made, and changing the meaning of cues. In this study, bilateral movements were categorised as high complexity, whereas unilateral movements were low complexity. When unilateral movements were included in high complexity lessons, participants were required to execute these movements with their non-dominant hand or foot. Additional requirements such as balance, the cognitive demand through external pacing of exercises (by the beat of music or being dictated by the pace of a game), and the application or variation of complex rules also put exercises into the high complexity category.

### Physical Education Lesson Content

The skill complexity in this intervention was controlled by careful skill selection for each intervention lesson. The low complexity lessons utilised exclusively simple, predominantly unilateral movements, avoided actions which place a demand on balance, and excluded the use of equipment requiring hand-eye/foot-eye coordination. In contrast, the high complexity lessons involved coordination of all four limbs simultaneously (aerobics) and required participants to manipulate equipment with their hands and feet in rapidly changing ways and environments (ball skills and games). To further increase the complexity, participants were required to use their non-dominant hand/foot for all unilateral skill practices.

*Condition 1.* Flexibility testing and simple obstacle course (low intensity, low complexity).

Warm-up: Walking pace moving around the PE space, walking with different stride lengths (giant steps, tiny steps, normal steps), followed by static stretching whilst being told what today's lesson will be.

Main activities: Sit-and-reach test in two groups. Participants are instructed to remain seated in a designated area until it is their turn. They are allowed to chat whilst waiting. After everyone has completed the test, extended static hamstring and posterior deltoid stretches are performed (all seated). Sit-and-reach test is repeated to see if stretching has successfully aided flexibility. Finally, participants complete a simple obstacle course (step over a bench, weave between cones) at walking pace, with a partner or small group.

Complexity avoided by: use of simple movements, no object manipulation, no specified use of non-dominant hand/foot.

Intensity maintained as low by: performing exercises in a seated posture or at no more than walking pace.

*Condition 2*. Juggling and bilateral ball-skills (low intensity, high complexity)

Warm-up: Club 100 – standing still or walking pace, using the non-dominant hand, throw a tennis ball about 1-2m into the air and catch again with the same hand. Repeat this ten times in a row without dropping the ball, then 20, 30, etc until you reach 100 in a row. Followed by static stretching whilst being told what today's lesson will be.

Main activities: Juggling (2 tennis balls) – after a teacher-demo, have a go at juggling two tennis balls, using the 'round-the-world' technique. Take turns with a partner to see how many times you can catch the ball in 30 seconds when juggling two tennis balls. Then, a series of bilateral ball skill drills (based on drills from the MFT (Rusch and Irrgang, no date) and the study by Budde *et al.* (2008)), with fast-paced switches between drills. With a partner, stand a few meters apart and face each other and:

One person has a football, the other a netball. Communicate with your partner to say 'go' and pass the football using the non-dominant foot, and netball using both hands, at the same time.

One person has a netball, the other has no equipment and turns their back to their partner. The participant with the netball calls their partner's name and passes the netball, their partner turns around and catches the netball.

Taking turns, dribbling a football between 'gates' of cones to score points, using the non-dominant foot.

Finally, participants completed more juggling practice and a 30-second re-test.

Complexity achieved by: fast-paced switches between drills, use of feet and hands or both hands at the same time or in quick succession, use of non-dominant hand/foot, object manipulation – multiple objects, reaction time.

Intensity maintained as low by: juggling practice done seated. Other drills are stationary or low-pace due to the demand for maintaining control, with the exception of the dribbling the only locomotion required is when control of the equipment is lost, and it needs to be retrieved.

*Condition 3.* Fitness testing (medium intensity, low complexity)

Warm-up: Exercises on the spot (jumping jacks, high knees, squats, spotty dogs) with a sprint on the spot on 'go', counting down from 10 to 1. Jogging around in a circle, doing different exercises as '1', '2', or '3' is called (jump up, touch ground, change direction). Followed by static stretching whilst being told what today's lesson will be.

Main activities: Timed 10m sprint (two or three times, depending on time), with two set-ups working at the same time to reduce waiting time. With a partner, counting for each other: one-minute sit-up test, and press-up test to failure. Whole class given time to practise standing long jump by themselves, then performed in groups of ¼ of the class with furthest three progressing to the 'final'.

Complexity avoided by: using simple and/or routine movements, no object manipulation, no bilateral coordination.

Intensity maintained as medium by: taking turns to balance intense exercise to achieve maximum scores in 10m sprint and sit-up/press-up tests, continuous practising of standing long-jump and limited waiting time for test by performing in large groups.

*Condition 4.* Team games (medium intensity, high complexity)

Warm-up: Rats and ravens – participants form two lines in front of the teacher, standing side-by-side with a person in the other line. Line 1 are the rats, line 2 are ravens. When 'rats' is called, the rats need to run to their 'safe zone', away from their raven partner, before their partner tags them, and vice versa. If they get 'home' without being tagged they score a point, otherwise their partner gets the point. Change the starting position (sitting, lying prone, crouching) after each round. Followed by static stretching whilst being told what today's lesson will be.

Main activities: Passing with non-dominant hand to a partner – catching with non-dominant hand. Starting stationary, then on the move down a straight line, then down a straight line with a defender in the middle, and finally in a small-sided game with free movement inside a designated court. The games were played as a tournament, with short games (2-3 minutes, depending on time available), and quick changes of opponents. In each round of the tournament, the way to score points was changed (based on increased complexity in the study by Schmidt *et al.* (2015b)); making five successful passes in your team, making five consecutive successful passes, passing to each team member consecutively.

Complexity achieved by: use of non-dominant hand, object manipulation, performing coordinated skills whilst moving, applying changing rules, combining object manipulation with spatial awareness (based on (Gallotta *et al.*, 2015), (Rusch and Irrgang, no date), and (Schmidt *et al.*, 2015b)).

Intensity maintained as medium by: balancing the competitive fast pace of the games with taking turns in the order of matches, a balance of stationary and moving drills, reduction in moving speed due to the need to maintain control of the object.

*Condition 5.* Sprinting games (high intensity, low complexity)

Warm-up: Truck and trailer – with a partner, stand behind each other. The person in front is the truck, the person behind the trailer. The truck runs around and the trailer follows, but the truck tries to lose the trailer whilst the trailer tries to stay close. On the whistle, freeze – and if there is more than

an arm's length distance between you the truck wins, otherwise the trailer wins. Switch roles after each round. Followed by static stretching whilst being told what today's lesson will be.

Main activities: Multi-Stage Fitness Test (beep test) to a maximum of level 10. Depending on the space available in the schools, a 15m or 20m test was performed. When participants failed to keep up with the pace of the test, they were encouraged to keep running on the side of the test set-up 'to encourage their classmates'. Once all participants completed the test, a brief practice of running (using arms to make you faster, long strides) was done, with all participants lined up next to each other so they could all work at the same time with no waiting to take turns. The remainder of the lesson consisted of multi-lane fast-paced 10m competitive sprints. After each sprint the participants were encouraged to jog back to the start line and quickly line up for their next sprint. Two four-lane set-ups were running at the same time, using the classroom assistant to supervise the second set-up. The final activity was a straight-line relay with tag/high-fives as a change-over.

Complexity avoided by: using simple, routine movements only (straight-line running, based on the studies by Best (2012) and Gallotta *et al.* (2015)), and avoiding object-manipulation (no batons for the relay).

Intensity maintained as high by: encouraging participants to remain active after completing the beep test. Fast-paced turn-around between sprints. During the brief running practice all participants were moving almost continuously and at the same time.

*Condition 6.* Aerobics (high intensity, high complexity)

Warm-up: Freestyle dancing to frequently changing music. Participants are encouraged to reflect the music's energy level in their dance style. Followed by static stretching whilst being told what today's lesson will be.

Main activities: Aerobics routines, copying the teacher. Learning each routine, building the complexity of the skills, then performing it continuously for the duration of an entire song. In the final 5 minutes participants created and performed their own sequence of up to four different four-beat moves, repeated for 1 minute continuously.

Complexity achieved by: aerobics moves incorporating simultaneous arm and leg movements in different directions (e.g. stepping sideways with arms going forwards). Keeping in time with the beat of the music (externally paced exercise). Turns during routines and needing to re-orient to perform the next part of the movement sequence in a different direction (all elements based on the study by Gallotta *et al.* (2015)).

Intensity maintained as high by: using fast-paced music, performing routines for the full duration of a song, allowing only short breaks between routines, incorporating movements with arms above shoulder-height, a high proportion of jumps in the routines, and having the teacher at the front of the class to model the sequences so that if children lost track they could immediately re-join.

Full lesson plans can be found in appendix G.

### 3.3.2 Dependent Variable: Behavioural Engagement

The primary outcome variable in this study is participants' observed on-task behaviour during classroom lessons after PE. This is used as a measure of their behavioural engagement. This study uses behavioural engagement as its outcome variable, as behaviour can be assessed through observations in a naturalistic environment (Fredricks *et al.*, 2004). In studies of classroom behaviour, the terms 'on-task', 'engagement', and 'attention' are used without harmony of the precise operationalisation, and the definitions and concepts vary between studies, depending on the research question(s) addressed. Common measures are time spent on-task, or off-task (Godwin *et al.*, 2013; Pellegrini *et al.*, 1995; Pellegrini and Davis, 1993), with off-task often further categorised. In this study, on-task behaviour is operationalised as 'displaying goal-directed and task-appropriate behaviours'. Studies using the Behaviour Observation of Students in School (BOSS) further differentiate between 'active' and 'passive' on-task behaviour, with active engagement requiring behaviours such as writing, asking questions, or speaking during a group discussion (Pearson, 2013). Passive engagement includes listening to the person who is speaking, quietly watching a video clip, or reading allocated material. As the potential for active engagement is often dictated by the teacher in the type of task they set the

student, and the research questions in this study centre around on-task vs. off-task behaviour rather than the qualitative type of engagement, the distinction between active and passive on-task behaviour was not applied in this study.

Many studies further categorise off-task behaviour, either by type of behaviour displayed (Heemskerk, 2015; Rock, 2005; Pellegrini and Davis, 1993), or the source of distraction (Godwin and Fisher, 2011). Godwin and Fisher (2011) focus on the source of the distraction, as this is a malleable factor. Heemskerk (2015) categorised 'off-task, independent', or 'off-task and disruptive to others', as social interaction was of interest in that study, which focused on play-deprivation, PA measures, and task-related social interaction (ibid.). Rock (2005) differentiated between active (fidgeting, unnecessary or excessive movement) and passive (staring, lack of participation, dozing off) off-task behaviours, in line with recommendation from Fredricks *et al.* (2004) regarding qualitative differences between withdrawn and disruptive off-task behaviours.

The coding applied to behavioural engagement in this study differentiates between active and passive off-task behaviour, based on conversations with primary school teachers who often describe the effects of PA as leaving children 'too tired' or 'too hyper' afterwards. Passive off-task behaviour is operationalised as 'an absence of goal-directed behaviour, and absence of physical or verbal actions', whereas active off-task behaviour is 'an absence of goal-directed behaviour, and presence of unnecessary physical or verbal actions'. Unlike the BOSS, no further differentiation was made for active off-task behaviour (verbal/motor).

### 3.3.3 Mediation Variables: Emotions

In this study, classroom-based behavioural manifestations of the emotion state effects of PA are investigated. More positive mood and emotions have been indicated as effects of PA participation (Chan *et al.*, 2018), and positive emotions have in turn been identified as precursors of improved engagement in the classroom (Fredricks *et al.*, 2004). Therefore, it is possible that emotional state may mediate the effects of PA on learning behaviour. PA participation has been related to lower

reports of depression (d=0.48) (Field, Diego and Sanders, 2001), and to improved mood through facilitating social interaction which increases feelings of positivity (Lubans *et al.*, 2016). Winter *et al.* (2007) found that PA improved ratings of positive mood for moderate as well as vigorous activity (d=0.59), but it did not affect ratings of negative mood. As well as their valence (positive or negative), emotions have an arousing value: activating or de-activating (Figure 1 (Inkinen *et al.*, 2013)). The combination of the valence and arousal affect behaviour in ways that can be prohibitive or facilitative to engagement, academic work and achievement (Fredricks *et al.*, 2004).

At three points during each classroom lesson and at the end of PE lessons, participants completed questionnaires on tablets. The questionnaire items about emotions provided information on the valence of the emotions (did they experience positive or negative emotions) as well as arousal intensity (how many stars they rated their emotions at), using 5-point Likert scales. The emotions included in the questionnaire were 'happy', 'excited', 'calm', 'relaxed', 'sad', 'bored', 'angry', and 'nervous'.

One participant in each class wore an Empatica E4 device, which recorded electro-dermal activity (EDA); an indicator of arousal. For these participants, EDA data and self-reports were planned to be combined to investigate the measured arousal from EDA alongside reported arousal and valence. Arousal features heavily in the relationship between cognition and exercise (Yerkes and Dodson, 1908; Tomporowski, 2003), and can be either emotional (psychological), or physiological (Schmidt and Wrisberg, 2004). It is not yet fully understood if these two forms of arousal have the same effect on cognition (Schmidt-Kassow *et al.*, 2013; Schmidt-Kassow *et al.*, 2014), and the current set-up of emotion self-reports and EDA recordings can be used to assess the contribution of emotion to arousal levels. Due to time constraints the E4 devices were not piloted for data analysis strategies, and technical difficulties meant that the EDA data could not be analysed at this point.

### 3.3.4 Control Variables

Additional background variables were collected to use as control variables in the multilevel linear models (MLM) and to investigate random effects at the participant level on the relationship between PA content and subsequent learning behaviour. These were collected through parental consent forms, teacher reports, participant self-reports, and by researchers recording physiological measures.

#### Attainment

Six of the eight teachers provided generic academic attainment scores for their pupils. These ratings on a 5-point scale, relative to participants' Age-Related Expectation (ARE) allow for easy comparison across age groups within the sample. Teachers rated the participants as '1. well below', '2. below', '3. at', '4. above', or '5. well above' ARE.

#### Height and Weight

Current literature suggests that people respond differently to PA, depending on their BMI (see section 2.3.7). Height and weight were recorded at the start of the intervention in order to calculate participants' BMI. Children removed their shoes and socks, and were then weighed using digital scales (Salter, model 9018S SV 3R), and measured with a measuring tape affixed to the wall. Their weight was recorded in kilograms, rounded to the nearest tenth of a kilogram, and their height in metres, rounded to the nearest half centimetre. BMI was calculated using the equation $BMI = \frac{weight\ (kg)}{height\ (m)^2}$. As BMI cut-points for weight status differ by age and gender, each participant was then assigned a BMI category: 'low BMI' (below 2nd percentile), 'low to healthy BMI (2nd to 25th percentile), 'healthy BMI' (25th to 75th percentile), 'above healthy BMI' (75th to 91st percentile), 'overweight' (91st to 98th percentile), or 'obese' (above 98th percentile). The allocation was made based on the participant's gender and age (rounded to the nearest half year), using the BMI curves from the Royal College of Paediatricians and Child Health (Royal College of Peadiatrics and Child Health, 2013).

### Age

The sample spans three year groups, with the oldest pupils approaching the start of puberty. To investigate if age is relevant to participants' response to PA, their date of birth was collected on parental consent forms and used to calculate their age in years and months on the first day of the intervention.

### Self-Reported Lesson Experiences

During each classroom and PE lesson, participants completed questionnaires, using 5-point Likert scales. Three times per classroom lesson they rated their enjoyment of the current task, level of tiredness, and perceived difficulty of the current task. At the end of PE lessons, they rated their enjoyment of the PE lesson, and how tiring they found the PE lesson. More details of the questionnaire can be found in section 3.4.2.

### Physical Activity Habits

Studies investigating fitness levels or exercise participation often split their sample into those who are 'lower fit' versus those who are 'higher fit' based on fitness testing, or 'less active' versus 'more active' based on reported exercise participation or activity data recording. In the present study, no standardised fitness testing was performed due to a combination of time restraints, class sizes and participant ages. Three different measures of PA habits were collected in this study: subjective measures of weekly sports participation frequency and duration through self-reports, and objective measures through 24h recordings of the GENEActiv devices.

On the final intervention day, participants completed a tablet-based questionnaire about their exercise habits. For each day of the week they recorded if they would normally have a PE lesson on that day (intervention not included), and if they would normally take part in any organised sports after school. If they did, they also recorded the duration of this sports session (for more details about the

questionnaire, see section 3.4.2). Field *et al.* (2001) used the sample mean of exercise participation to allocate participants to the less or more active groups and found that those who participated in more exercise than average had more positive relationships, reported less feelings of depression, and had higher academic achievement. In this study, participants were allocated to two groups based on their scores: less than average participation and more than average participation. They received two allocations; one based on the number of times per week they exercised, and one based on the duration in minutes they exercised.

Coe *et al.* (2006) applied the 'Healthy People 2010' guidelines and split their sample into those who did versus those who did not meet the guidelines. They found no effects on academic performance based on PE lesson participation, but that those meeting the guidelines outperformed those who did not on all measures of academic achievement (ibid.). In this study, participants wore accelerometery devices for 24h leading up to each intervention period (up to six times 24h per participant). The mean number of minutes of MVPA they accumulated over 24h (excluding the PE lesson) was compared to the current government guidelines of at least 60 minutes of MVPA per day (Chief Medical Officers, 2011). Participants were then allocated to one of two groups: meeting MVPA guidelines, or not meeting MVPA guidelines. Finally, if they achieved the government guidelines on at least 75% of recorded days they were assigned to the 'active child' group.

## 3.4 Data Collection Instruments

### 3.4.1 Observations

Observations were carried out by two observers who were present in the classroom during lessons. Momentary time sampling was used as the method for observations. Although continuous recording methods would be the most precise (Meany-Daboul *et al.*, 2007), this was not feasible in this study as more than one student needed to be observed in each classroom. Time limitations and the additional difficulties involved in obtaining ethics approval and school/parental consent for the use of video

recordings in schools in the UK (Jewitt, 2012) prevented the use of video to then code retrospectively for several students in the same classroom, using a continuous recording method.

Further barriers to the use of video recording were the additional burden on teachers and participants for wearing sound recording equipment. This, along with the presence of the camera in the room, may also influence behaviour, as it reminds the participants that their actions are being recorded (ibid.). However, the influence of the presence of the camera needs to be weighed up against the presence of the researchers for observing participants. Another drawback of watching a video versus physically being in the room, is the possible loss of the essence of the events taking place (ibid.). Moreover, live observations in the classroom allow the observer to make small adjustments to their position to see the face or body of their observation target when it would otherwise be obstructed due to movements of the target or other people in the classroom. This would not be possible when watching a video recording, as video does not allow for peripheral vision, is stationary, and is limited by the lens angle, thereby increasing the potential for missing data (Jewitt, 2012).

In momentary time sampling, a predetermined recording interval is chosen, and at the end of each interval the observer looks at their target person for a few seconds to determine their current behaviour, which is then recorded (Meany-Daboul *et al.*, 2007). This method can lead to under- or overestimation of behaviour, especially underestimation of low-frequency behaviours, as only a fraction of the total observation period is actually observed (Zakszeski, Hojnoski and Wood, 2017). However, as the observation intervals are predetermined, the under- or overestimation occurs in an unbiased way (Meany-Daboul *et al.*, 2007). Existing momentary time sampling tools use observation periods of around two to three seconds, with intervals between observations at two to 60 seconds (Pearson, 2019; 2013). Two seconds is recommended to get the closest estimations to the actual occurrence frequency of the behaviour, whereas 60 seconds is more suitable for the observation of high-frequency behaviour (Pearson, 2013; Zakszeski *et al.*, 2017). In this study, the observation intervals were set at 30 seconds, so that lower-frequency behaviours such as active and passive off-task behaviour would be less likely to be underestimated. Zakszeski *et al.* (2017) compared

observation intervals of five, ten, fifteen, 30, and 60 seconds, and found that observations using intervals of five to fifteen seconds were most similar to continuous observation methods. However, they recommend using an interval that is most feasible for the study at hand, and especially when multiple participants are observed longer intervals may be more appropriate (ibid.). Six participants were observed by each researcher in this study, therefore a 30 second interval gave observation periods of five seconds. During this time, the researcher would look at the target participant, decide their behaviour category, and log this on the specially designed app. Pilot studies were conducted using observation protocols with five second observation periods (see section 3.6), and this duration was found adequate to decide and log the behaviour category without feeling pressured for time. Especially with the use of the app rather than the paper and pencil method, five seconds was ample time.

Observations were carried out by the researcher and a research assistant, using a tailored protocol consisting of four behaviour categories and five task categories (appendix C). Participants were rated as 'on-task' if they displayed goal-directed behaviour to carrying out the task that was instructed by the teacher. If they were not on-task, a distinction was made between 'passive off-task' or 'active off-task' to assess if the PE lesson content affected participants' energy levels in the lesson after PE. Passive off-task behaviours included yawning, stretching, staring into space, and daydreaming, whilst active off-task behaviours included speaking if instructed to work in silence, and the execution of unnecessary movement that interfered with task performance. The most important determinant of off-task behaviour was the absence of goal-directed behaviour. Therefore, if children performed unnecessary movement whilst carrying out their task, e.g. tapping their foot/pencil whilst reading a book, or they yawned whist writing, they were rated to be 'on-task'. This is in line with how behaviour is categorised using the BOSS (Pearson, 2013). An additional category of behaviour was 'other' for moments when participants were not on-task, but their behaviour was driven by a biological need, such as needing a drink, or if the rater could not identify the behaviour as either on-task or off-task.

The specification for the different behaviour categories was found to be dependent on the task and teacher's expectations by Heemskerk (2015). The full observation protocol takes this into account and defines the behaviour categories dependent on five different task types: independent, paired work, group work, teacher-led, or other (appendix C). Due to this dependency, it was vital for the raters to pay close attention to the teacher's instructions during observations as well as rating the behaviour, and task type was also recorded with each observation. Independent work was any task where the participant was expected to carry out a task without assistance, and without interacting with others. Paired work allowed for communication with one other pupil, whereas group work was any task where three or more pupils were working to complete the same piece of work, or discussion between 3 or more pupils was allowed or expected. Teacher-led work consisted of the class, or a group of pupils, listening to the teacher or one allocated other person (such as a teaching assistant, or a pupil who had been allocated to ask or answer a question). On occasion, pupils were asked to watch a television clip. This was considered a teacher-led task as pupils were expected to sit and listen, with the television effectively taking on the role of the teacher. The final category, 'other', covered non-learning task activities such as transitions between tasks, the filling out of research questionnaires, and waiting for resources to be handed out. For the majority of occasions, the 'other' category still consisted of a task, and as such had behaviour expectations. Participants were therefore still rated on their behaviour during the 'other' tasks. Task type was also recorded during observations, as the incidence of on-task behaviour has been found to differ depending on task type (Heemskerk, 2015; Godwin *et al.*, 2013)

Observations took place during classroom lessons before and after PE lessons. Each participant was observed every 30 seconds, with six participants per observer (12 per classroom). To allow for observations immediately before and after PE lesson, without the interruption of break or lunch, the research was always carried out in the afternoon. Observations before PE started when the pupils returned from lunch, and after PE lasted until home time (or a few minutes before, depending on what time school finished). Pellegrini *et al.* (1995) observed behaviour for 20 minutes pre- and post-break.

The observed periods in this study were 25 to 30 minutes long. Initially, 30 minutes was decided upon as this would provide more opportunity to see if transient effects of PE 'wear off' over the course of the subsequent lesson. However, with restrictions posed by the time available between the end of lunch and home time, 30-minute observation periods could not be facilitated in all schools, and in schools 2, 3 and 4 they were adjusted to 25 minutes. This should still be sufficient, based on findings by (Chang *et al.*, 2012) that the greatest effects of acute PA occur eleven to twenty minutes after cessation of the exercise.

Two raters carried out observations. Four rounds of inter-rater reliability testing were carried out to ensure consistent application of the observation protocol (Cohen's Kappa=0.8). More details of the inter-rater reliability testing can be found in section 3.7.




*Figure 11. Observations app interface.*

An app was designed to collect the observation data for this study, and programmed using Android Studio (Google LLC, 2017). This was chosen over the traditional pen and paper method to facilitate more frequent observations (every five seconds) without the need for the observers to flip pages and manage a clipboard with the paper, as well as a stopwatch or timing device. As the teacher's instructions were vital to take into account when deciding on the behaviour category and task type, a pre-recorded beep every 5 seconds via an earpiece was not deemed viable.

The app consisted of a single page with a stopwatch, and buttons for each participant (numbered one to six to maintain anonymity) alongside buttons for behaviour and task categories (Figure 11). Date and time stamped data for each observation was saved as a comma separated file on the device and transferred to the researcher's laptop. The order in which

participants were observed was recorded separately and SPSS 24.0 (IBM Corporation, 2016) was used to replace the participant number recorded by the app with the participant ID codes.

### 3.4.2 Questionnaires

#### Lesson Experience Sampling

Momentary self-reporting on tablets was completed up to three times during classroom lessons in which participants were observed (at five, 12.5 and twenty minutes from the start, representing the start, middle and end of the lesson), and once at the end of PE lessons. Questionnaires focused on collecting data about participants' lesson experience to use as control variables (enjoyment, tiredness, task difficulty, see section 3.3.4), as well as the emotion ratings discussed in section above3.3.3. Figure 12 shows the user interface of the questionnaire, the full questionnaire is available in appendix H. All ratings were completed by tapping on a star-bar, representing a 5-point Likert scale. The emotion items were based on the Positive and Negative Affect Schedule (PANAS) which uses 5-point scale ratings (Watson, 1988) and is a widely-used tool for assessing emotional state. Items were worded to reflect the momentary nature of the assessment, rather than trait affect (Merz and Roesch, 2011).

As observations of behaviour differentiated between passive and active off-task behaviours, participants were asked to report on their level of tiredness. Tiredness can be further linked to PA habits and sleep patterns (Lubans *et al.*, 2016), as well as relating to emotional arousal, and possible physical arousal from acute PA. Perceived task difficulty has been related to learning behaviour and motivation (Malmberg *et al.*, 2013a; Malmberg and Little, 2007) and it was suggested by Martin *et al.* (2015) that future studies of the stability of learning behaviour and motivation should include perceived difficulty in the design. Task enjoyment is of interest in the present study as enjoyment of learning is related to intrinsic motivation, and facilitates positive learning behaviours and engagement (Malmberg, Woolgar and Martin, 2013b).

To reduce the burden on participants from the repeated answering of questionnaires (up to 7 times per afternoon), the questionnaires were kept as short as possible. It was decided to ask for ratings of only two emotions from each arousal-valence quadrant. This did lead to a trade-off between reduced measurement validity due to only having two observed variables for each latent emotion factor on the one hand, and increased participant compliance due to a decreased burden on the other hand (Dunton *et al.*, 2014). The potential burden from repeated questionnaires was further decreased by the use of tablets instead of pen and paper questionnaires. The specifically designed questionnaire app used emojis (Rane, 2017) and made the questionnaire more accessible to young children by reducing the amount of reading required as well as shortening the time required to complete the questionnaire.

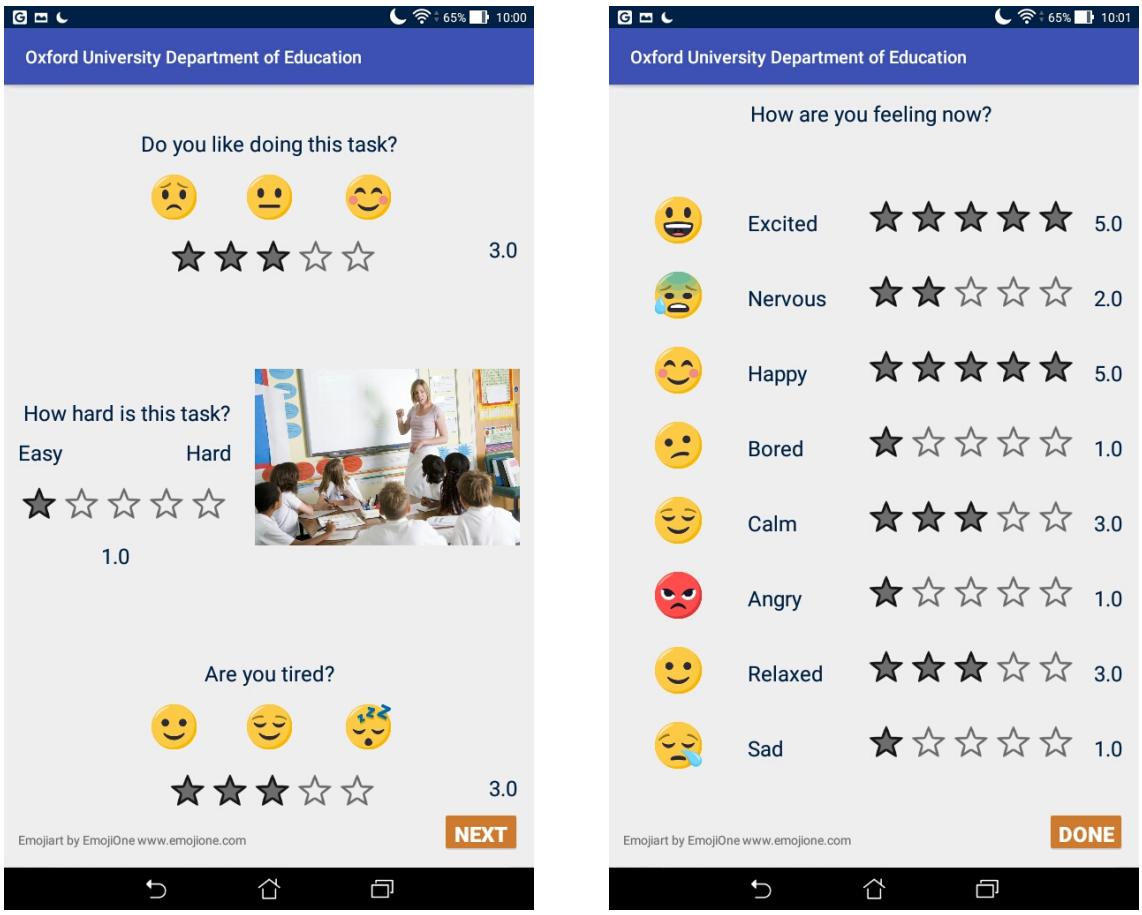



*Figure 12. Questionnaire app interface.*

### Physical Activity Habits Questionnaire

On the final day of intervention participation, participants completed the trait PA questionnaire. The trait PA questionnaire was also app-based and used the same type of emojis as the learning experiences questionnaire. It consisted of two questions per day for Monday to Friday, and one question for the weekend days. For weekdays, participants were asked if they would normally take part in PE on this day, and if they did any sports after school. If so, they were asked which sport and how many minutes they did these sports for. For each day they could log a maximum of two after-school sports. For weekend days, participants were only asked about sports participation, and again logged which sport and how many minutes they took part for.

## 3.5 Sampling Strategy

GPower software (Faul *et al.*, 2017) was used for a-priori sample size calculations. This calculation was based on the repeated measures element of the research design. To take into account the multi-level structure of the sample (children nested in classes, classes nested in schools), De Jong, Moerbeek and van der Leeden (2010) established the 'design-effect', an additional factor by which to multiply the suggested sample size, based on the intra-class correlation (ICC). The ICC in this study will be at the level of the classroom, and thus (in UK primary schools) the teacher. The Learning Every Lesson study by Malmberg *et al.* (2015) found an ICC of 0.10 between teachers for students' autonomous motivation, and this is used as an approximation of the ICC for this study (De Jong *et al.*, 2010). This led to a total sample size estimation of 66, but due to an expected drop-out of 30% the intended sample size for recruitment was 96. The full calculations can be found in appendix F.

After achieving CUREC approval (appendix I), participants were recruited from four primary schools (eight classes) in Oxfordshire. Initially, schools with staff members known to the researcher were approached. This provided a convenience sample (Sedgwick, 2013) of five classes in two schools. Although this type of sampling reduces the external validity of the study by preventing generalisability

outside of the sample (Sedgwick, 2013; Rothwell, 2005), due to the difficulties of obtaining access to schools it was chosen nonetheless. Moreover, the schools considered for inclusion in the study vary greatly in the characteristics of their catchment area. For the next round of sampling, all schools in Oxfordshire were contacted via email to invite them to participate. Another two schools (three classes) were recruited this way. Contact was made with the head teachers in each school, and consent obtained from them to carry out the research in their school.

Purposive sampling then determined the eventual sample (Coolican, 2009), to ensure a roughly even gender split and check suitability of the participants. Classes in years 3 to 6 were included in the study design and pupils with informed parental consent were screened for criterion compliance; physically fit and able to participate in all PE lessons, no ADHD/ADD, epilepsy, or statemented learning difficulties. ADHD/ADD and epilepsy were important exclusion criteria as they are known to affect behavioural engagement or neurological processes (Pimentel, Tojal and Morgado, 2015; Jarrett *et al.*, 1998). Especially where medication is involved, the scheduling of such medication may affect the behaviour and attentiveness of the participants (Jarrett *et al.*, 1998). From those with parental consent and compliant to inclusion criteria, a sample of twelve participants plus two spares was selected in each class.

### 3.5.1 Sample Characteristics

#### School Characteristics

The four schools taking part in the study are very diverse; both in size (Number on Roll between 117 and 401) and the catchment area within Oxford(shire) they are located in.

School 1 is an academy converter, already partly integrated with the local academy group of schools. It is located about half an hour outside of Oxford. It is the smallest of the participating schools, with just four classes in the school. The year 5 pupils who took part in the study were in a combined class with the year above (n in class 33). Pupils normally take part in two PE lessons per week. Some

lessons take place on the school's field, or in the (very small) hall. Once a week the older pupils have PE in the hall at the nearby community centre. All PE lessons are taught by the classroom teachers, and these also run after school clubs alongside external coaches.

School 2 is a voluntary aided Church of England school in Oxford. This is one of the two largest schools in the sample, with double form entry. Two year 4 classes (n in class 29 and 28) and two year 5 classes (n in class 30 both classes) took part in the study. Pupils at this school have two PE lessons per week, but for some terms one of these lessons is substituted with 'outdoor learning' in a nearby nature reserve. The school has a large field and courts, as well as a full-size sports hall. All PE lessons are taught by the classroom teachers. A large number of after school sports clubs is offered by the teachers, supplemented further by external coaches.

School 3 is an academy in the outskirts of Oxford, and also one of the largest schools with a double form entry. Both year 4 classes participated in the study (n in class 26 both classes). They normally have two PE lessons per week, either in the hall, or on the generous field or courts outside. Up until year 4 all PE lessons are taught by the classroom teachers. A PE specialist teaches the year 5 and 6 PE lessons, as well as several after school clubs. Once a term, those children who have accumulated a minimum of 50 reward points throughout the term (in all lessons and wider school life) take part in an additional 'reward PE session' with the PE specialist.

School 4 is a voluntary aided Church of England school in Oxford with a single form entry. The year 3 class took part in the study (n in class 28). Pupils are taught two PE lessons per week, taking place in the school hall or outside on the field or playground. Across all years, some PE lessons are taught by classroom teachers and others by a PE specialist. Classroom teachers and the PE specialist also offer after school clubs.

### Participant characteristics

A total of 110 participants were recruited across the four schools. After criterion compliance screening, two were excluded on medical grounds (one with a steroid inhaler and one diagnosed with

epilepsy during the intervention). All of the remaining 108 participants (43 male/65 female, $M_{age}$ 9y 4m, SD 7.5m, age range 7y 11m – 10y 6m) answered questionnaires, and 101 (41 male/60 female, $M_{age}$ 9y 4m, SD: 7.5m, age range 7y 11m – 10y 5m) wore accelerometery devices and were observed.

## 3.6 Pilot Studies

Two projects have served as pilot studies for this study: my MSc dissertation research (Heemskerk, 2015), and the Activity, Rest and Learning Experiences (ARLE) study (Malmberg, de Vos and Saunders, 2016). The MSc research used observations and self-reports for participants in years 4 and 5, in classroom lessons before and after bouts of PA. Participants in the study wore basic pedometer bracelets. In the ARLE project, pupils in years 4 and 6 wore GENEActiv devices 24/7 for up to 2 weeks, and reported on their learning experiences during this period using tablets. The pilot studies provided valuable opportunities to try different data collection devices.

The IV in this study is the type of PA performed in the PE lesson. The intensity of the exercise is measured by physiological data from accelerometery. This study will use the same accelerometery devices as the ARLE (GENEActiv), as well as Empatica E4 bracelets, rather than pedometers. The accelerometery devices provide more reliable movement data than pedometers, and can be used to differentiate between sedentary, light, moderate, and vigorous exercise. Additionally, they do not have a display, and are therefore less likely than pedometers to prompt the participants to participate in PA they normally would not.

The main DV in this study is behavioural engagement, which was also the DV in previous MSc work. A similar observation protocol was used, and 5 seconds was found adequate time to observe and record behavioural engagement and task type. The use of an app for observations rather than pen and paper was prompted by the issues with flipping pages on the recording sheets, juggling the clipboard as well as a stopwatch, and the time spent on and potential for error from manually entering the data into a computer programme. As part of the ARLE research, the protocol was tested for interrater

reliability (see section 3.7). Adaptations have been made to the classification of off-task behaviour (active/passive instead of independent/with others), and clearer, more detailed categorisations produced. Further interrater reliability testing was carried out in November 2017, February 2018, and May 2018.

Emotional state was investigated as possible mediator of the effects of PA on behavioural engagement. This was assessed through self-rating scales. Both the ARLE project and MSc work used repeated self-rating questionnaires during lessons. Although these did not take long to complete, they did create some disruption to lesson flow. To minimise disruption, it is essential to keep the questionnaires short, with items that are simple and straightforward for children to answer. Different devices were used in the two studies (handheld computers and tablets), which the children enjoyed working with. The tablets offered greater scope for personalisation of the questionnaire, and the use of graphics. Graphics can reduce the amount of reading required and make the questionnaire more accessible to younger children and those with lower reading attainment. The ARLE study highlighted the need for this, as some of the year 4 pupils required help from the teacher in reading and answering the questions. This is especially pertinent to this study, as the target sample extends to younger children (seven-year-olds) and is larger – and thus more likely to include a range of academic abilities. The use of graphics will also reduce the time required to complete the questionnaire, therefore impacting less on the flow of the lesson. The app was created to function without the need for an internet connection, so that there was no need for reliance on school WiFi networks and the associated potential for loss of data in the event of poor connectivity.

## 3.7 Reliability

This study of behavioural engagement in a naturalistic setting is highly susceptible to contextual influences, reducing its reproducibility and reliability (Coolican, 2009). To improve reliability and

minimise fluctuations in contextual factors, several procedures were added to the protocol concerning the PE lesson delivery, classroom lesson content, and inter-rater reliability.

Firstly, the PE intervention was delivered by the researcher in all schools, so that the content and delivery of the PE lessons remained constant. Inevitably, the resources and physical space the lessons took place in varied somewhat between schools and with weather conditions. Whenever possible lessons were taught outside, but rain, snow, and freezing conditions made this unattainable at times.

Secondly, the classroom lessons before and after the PE lesson were taught by the usual class teacher and they were asked to keep lesson content constant across all six conditions. However, teacher compliance issues and wider school factors such as test weeks and fire drills meant that some lessons were changed at the last minute and without prior notice to the researchers. To address this it was decided that, rather than entirely excluding from the analyses those lessons which did not fully comply with the protocol, additional control variables would be used, such as self-reported task enjoyment. Within each intervention day, the lesson before and after PE was taught by the same classroom teacher in all cases. This study aims to detect effects of PE lesson content in a naturalistic setting; if these effects are not detectable above the amount of 'noise' stemming from the normal goings-on in classrooms and the fact that things in schools do not always go to plan, they will not be of value to teachers in their day-to-day practices. Therefore, task type, logged as part of the observation protocol, and self-reported task enjoyment from questionnaires were included as control variables in analyses.

Thirdly, inter-rater reliability testing was performed to increase reliability of the observations (Mabry, 2008). The preliminary version of the observation protocol was tested during pilot studies and led to Cohen's kappa of $\kappa$=0.72. The protocol was adapted with improved and more precise operationalisation of each category of behavioural engagement (Yin, 2013) (appendix C), and was tested before each round of observations, achieving $\kappa$=0.8 (SE=0.04, 95% CI=0.73-0.87). The kappa statistic was chosen rather than the ICC or percentage agreement, as it takes into account the likelihood of raters agreeing due to chance (McHugh, 2012). Especially as the on-task category was

expected to occur for around 75% of observations (Godwin and Fisher, 2011), and the occurrence of task type was not expected to be evenly distributed across observations either (Heemskerk, 2015), it was important to control for chance in the interrater reliability testing to avoid erroneously high agreement scores.

The kappa statistic ranges from -1 to 1, with 0 representing agreement resulting purely from chance, and 1 being perfect agreement between raters (McHugh, 2012). For the calculation of $\kappa$, both the actual agreement between raters and the agreement expected by chance are of interest. Table 1 gives example data for an observation period consisting of a total of 123 observations and including four different categories of behaviour. Category one is far more prevalent in these observations than any other category, giving high values for column 1 (c1) and row 1 (r1).

*Table 1. Example data for calculation of chance agreement of raters.*

| | | Rater 1 | | | | |
|---|---|---|---|---|---|---|
| | | On-task | Off-task passive | Off-task active | Other | Row total: |
| Rater 2 | On-task | 39 | 4 | 3 | 5 | 51 (r1) |
| | Off-task passive | 3 | 12 | 2 | 6 | 23 (r2) |
| | Off-task active | 5 | 2 | 19 | 2 | 28 (r3) |
| | Other | 6 | 4 | 1 | 10 | 21 (r4) |
| | Column total: | 53 (c1) | 22 (c2) | 25 (c3) | 23 (c4) | 123 observations |

The expected agreement between raters based on chance can be calculated using Equation 1, whereby n is the total number of observations. The actual agreement (as a raw percentage) is calculated using Equation 2.

$$Pr_{expected\ by\ chance} = \frac{\frac{c1 * r1}{n} + \frac{c2 * r2}{n} + \frac{c3 * r3}{n} + \frac{c4 * r4}{n}}{n}$$

*Equation 1. Agreement expected by chance (McHugh, 2012).*

$$Pr_{actual} = \frac{agreement\ category\ 1 + agreement\ category\ 2 + \cdots}{n}$$

*Equation 2. Actual agreement as a raw percentage (McHugh, 2012).*

In this example, the agreement expected by chance is:

$$Pr_{expected\ by\ chance} = \frac{\frac{53 * 51}{123} + \frac{22 * 23}{123} + \frac{25 * 28}{123} + \frac{23 * 21}{123}}{123} = 0.29$$

And the actual agreement is:

$$Pr_{actual} = \frac{39 + 12 + 19 + 10}{123} = 0.65$$

These values are then used to calculate the Kappa statistic, using Equation 3.

$$\kappa = \frac{Pr_{actual} - Pr_{expected\ by\ chance}}{1 - Pr_{expected\ by\ chance}}$$

*Equation 3. Cohen's Kappa (McHugh, 2012).*

In this example, $\kappa = \frac{0.65-0.29}{1-0.29} = 0.51$, which is considerably lower than the raw percentage agreement of 0.65. It is clear that it is harder to achieve a high value for Cohen's Kappa than it is to achieve a high raw percentage agreement. Especially when one behaviour category is particularly prevalent, and the value for $Pr_{expected\ by\ chance}$ is correspondingly high (Equation 1), a high value for $\kappa$ is harder to achieve (Equation 3). Next, to determine the 95% confidence interval, the standard error was calculated using Equation 4, and then utilised for computing the 95% confidence interval (Equation 5).

$$SE_{\kappa} = \sqrt{\frac{actual\ agreement\ (1 - actual\ agreement)}{n(1 - expected\ agreement\ by\ chance)^2}}$$

*Equation 4. Standard error of Cohen's Kappa (McHugh, 2012).*

$$95\%\ CI = \kappa +/- 1.96 * SE$$

*Equation 5. 95% Confidence Interval for Cohen's Kappa (McHugh, 2012).*

In this example, the standard error of the Kappa statistic is:

$$SE_{\kappa} = \sqrt{\frac{0.65\ (1 - 0.65)}{n(1 - 0.29)^2}} = 0.06$$

Which makes the 95% confidence interval 0.39-0.63:

$$95\%\ CI\ lower\ limit = 0.51 - 1.96 * 0.06 = 0.39$$
$$95\%\ CI\ upper\ limit = 0.51 + 1.96 * 0.06 = 0.63$$

Finally, to account for the variation in participants' fitness levels, and thus the perception of PE lesson intensity, the objective measure of PA from accelerometery was complemented with subjective

measures of exertion. This allowed participants' perceived physical exertion during the PE lesson and their resulting feelings of tiredness during the subsequent classroom lesson to be controlled for.

## 3.8 Validity

Several measures were integrated in the study design to ensure good construct validity. Firstly, the study was single-blind (Coolican, 2009). The classroom teacher was not aware of the study hypotheses, or the order of PE conditions the class participated in. Rosenthal and Jacobson (1968) found that the teacher's awareness, even if not explicit, has a significant effect on their actions and pupils' outcomes. A double-blind design could not be achieved in this study, as the researcher delivered the intervention lessons as well observing the participants in their lessons (Coolican, 2009). This increases the possibility that the observations will be less objective and is an issue the researcher must be aware of. Inter-rater reliability checks were performed for each block of observations to guard against the loss of objectivity (Mabry, 2008).

To ensure the differences in behavioural engagement observed can be attributed to the qualitative differences in PE intervention lesson content and not to classroom lesson content, classroom lessons were kept equivalent across the conditions, and control variables were collected for each participant (Harris, 2010). To prevent order effects of the six PE lessons, each group took part in the PE lessons in a randomly allocated order. Additionally, the time between PE lessons was at least 6 days, which is sufficient for any acute effects of the previous lesson to have 'worn off' before the next observation day (Coolican, 2009; Hoetting *et al.*, 2016; Schmidt-Kassow *et al.*, 2013). In some cases, there were two or more weeks between lessons due to school holidays or snow closure.

To minimise the effects of participants' awareness of being watched, thus altering their behaviour (Coolican, 2009), the researchers only took brief glances at the participants during observations, otherwise keeping their eyes on their tablet. The observations app included a stopwatch, so that researchers did not need to fix their gaze on any other objects.

## 3.9 Ethics

Ethical approval was granted by the Central University Research Ethics Committee (CUREC) at the University of Oxford, evidenced in appendix I. CUREC protocols 15 and 25 are applicable to this study; although physiological measures are collected with accelerometers, this is considered a non-invasive method and therefore does not necessitate medical approval.

This is an opt-in study. First, permission was sought from the schools' head teacher. Next, pupils were informed about the study during an afternoon visit from the researcher. Parent information was provided via a leaflet with the study details, including procedures for raising a complaint. Parents were required to submit a signed consent slip, and only those pupils with parental consent were considered for inclusion in the study. Finally, as suggested in protocol 25, participants were asked for assent. Those pupils who did not opt in took part as normal in both PE and classroom lessons, but without wearing accelerometers, filling out questionnaires, or being observed.

One of the IVs in this study was expected to have a curvilinear relationship with behavioural engagement, meaning that for some of the conditions involved in the intervention the subsequent behavioural engagement may be negatively affected (Covassin *et al.*, 2007). Moreover, individual characteristics such as fitness level may lead to curve-shifting (Schmidt and Wrisberg, 2004), and more than one lesson could consequently be affected by this (section 2.3.7). However, all of the activities in the PE intervention are those which are normally expected to be taught to pupils as part of the PE national curriculum (Department for Education, 2013) – the only difference being that now the subsequent lesson was analysed for behavioural engagement.

## 3.10 Analysis

### 3.10.1 Data Cleaning and Variable Computation

#### Observation Data

Observation data was removed if children were in a different setting (such as interventions or music lessons), or if both the task and behaviour were rated 'other'. The remaining observations were used only if a minimum of 40 observations were recorded during the lesson (5.6% of lessons discarded), leaving 51,640 observations. Behaviour variables were recoded into four binary variables (on-task, passive off-task, active off-task, other) with 0 indicating this behaviour was not displayed at the time of the observation, and 1 that this behaviour was displayed at the time of the observation. This created a dataset in long format (Table 2). Aggregate scores of behaviour were created per participant per lesson and used to create a new dataset with weeks nested in participants, and mean pre- and post-PE values for each variable for each week (Table 3). This wide dataset was used for the SEM. Additional aggregations were created for mean scores for each participant across all lessons in the intervention, and for three segments within each lesson (start, middle, and end). Segments were defined in the same way as for merging with the questionnaire answers (see section 3.10.2).

*Table 2. Data structure in long format (raw data, used for MLM).*

| ID | Week | Lesson | Obs nr | On-task | Passive off-task | Active off-task | Other |
|---|---|---|---|---|---|---|---|
| 1 | 1 | Pre | 1 | 1 | 0 | 0 | 0 |
| 1 | 1 | … | … | | | | |
| 1 | 1 | Pre | 50 | 0 | 0 | 1 | 0 |
| 1 | 1 | Post | 1 | 0 | 0 | 0 | 1 |
| 1 | 1 | … | … | | | | |
| 1 | 1 | Post | 50 | 0 | 1 | 0 | 0 |
| 1 | … | | | | | | |
| 1 | 6 | Pre | 1 | | | | |
| 1 | 6 | … | … | | | | |
| 1 | 6 | Pre | 50 | | | | |
| 1 | 6 | Post | 1 | | | | |
| 1 | 6 | … | … | | | | |
| 1 | 6 | Post | 50 | | | | |

| | | | |
|---|---|---|---|
| 2 | 1 | Pre | 1 |
| 2 | 1 | … | … |
| 2 | 1 | Pre | 50 |
| 2 | 1 | Post | 1 |
| 2 | 1 | … | … |
| 2 | 1 | Post | 50 |

*Table 3. Data structure in wide format (aggregated data, used for SEM).*

| ID | Week | Condition | On-task pre | On-task post | Passive off-task pre | Passive off-task post |
|---|---|---|---|---|---|---|
| 1 | 1 | 4 | 0.74 | 0.76 | 0.03 | 0.02 |
| 1 | 2 | 3 | 0.72 | 0.71 | 0.08 | 0.09 |
| 1 | … | | | | | |
| 1 | 6 | 1 | 0.73 | 0.78 | 0.06 | 0.02 |
| 2 | 1 | 4 | 0.71 | 0.71 | 0.11 | 0.12 |
| 2 | 2 | 3 | 0.74 | 0.77 | 0.09 | 0.07 |
| 2 | … | | | | | |
| 2 | 6 | 1 | 0.74 | 0.75 | 0.08 | 0.05 |

### Questionnaire Data

*Lesson Experience Sampling*. Questionnaire answers were disregarded if the answers to all questions were missing (six occurrences). Another 27 questionnaires had no emotion ratings, but did rate the task enjoyment, difficulty, and participant tiredness. These ratings were maintained, and the emotion items coded 'missing'. Additionally, if all emotion ratings in one questionnaire were the same, the emotion ratings were changed to 'missing' and the task ratings were maintained (110 occurrences). Across self-report items, missing data accounted for between 1.2% and 16% of all reports. More details on missing data from self-reports can be found in section 4.4.1.

*Physical Activity Habits*. PA questionnaires were screened for sports participation duration, and weekday activities were capped at 90 minutes per activity, with a total cap of 120 minutes per day. On weekends, a total cap of 240 minutes was applied (120 minutes per day). Any sports duration beyond this cap was deemed to be an outlier. For each participant, all sports participation minutes were added up to provide a weekly duration in minutes. Additionally, they were given a binary score for sports participation each day (0=did not participate in sports today, 1=did participate in sports today), and these were also added up across the week. As the weekend was assessed as a one entry,

scores ranged between zero and six. For both duration and frequency of participation, each participant was given a rating of '0' if they were below the sample mean, or '1' if they were at or above the sample mean.

### Accelerometery Data

Data from the GENEActiv accelerometer devices was processed with the GGIR package for R, version 1.10-1 (Migueles *et al.*, 2019; Van Hees *et al.*, 2014; Sabia *et al.*, 2014) and outputs were saved as comma separated files for reading into other software. As part of the processing, data was calibrated and only those sections of the recordings with valid movement data were retained.

GGIR was programmed to calculate values for each 24h period, as well as for sub-sections of each recording: end of school until 8pm, 8pm until midnight, midnight until 7am, 7am until the start of school, start of school until the start of lunch, lunch break, end of lunch until the start of PE, PE lesson, end of PE until the end of school. Only the 24h recordings containing at least 8h of valid data during the daytime (7am until 8pm), and PE recordings containing at least 30 minutes of valid data, were used in analyses. It was decided that including shorter recordings would give an inaccurate and unfair representation of participants' total movement time.

Data was condensed into 1-second epochs by the GGIR programme. Euclidean Norm Minus One (ENMO) cut-points were manually specified for LPA, MPA, and VPA based on a validation study by Fairclough *et al.* (2016) and research by Lonsdale and colleagues in the iPLAY study with primary school children in Australia (Sanders, personal communication, 27 March 2019). Minutes spent at each intensity were calculated by adding each second recorded at a given intensity, over the duration of each recording segment. More details on the ENMO variable and validation of the GENEActiv devices can be found in appendix B.

Difficulties in data processing for the Empatica E4 devices led to the movement data of seven participants being discarded. As the E4 devices' sampling frequency is 32Hz, and the GENEActiv devices sample at 100Hz, different bandwidth filters would be needed for the analysis. Moreover, to

the best of my knowledge, no validation papers have been published comparing the GENEActiv and E4 devices. These two factors lead to doubts about the consistency of the measures collected with these two different devices, and thus the reliability of any comparisons in cut-points from the data. It was decided that it would be preferable to exclude the Empatica E4 data from the analyses.

### 3.10.2 Merging Data from Various Sources

Data from observations and questionnaires was collected at different sampling frequencies. In each 25-minute observation period, questionnaires were completed at five minutes from the start, 12.5 minutes from the start, and twenty minutes from the start, to reflect the starter, main body, and plenary of the lesson. Participants would thus report on their starter task, and twice on their main task. Analyses of the MSc data compared models which merged observation and questionnaire data in four different ways. A 'retrospective model' merged the questionnaires to observations taking place before the questionnaire was completed, a 'predictive model' merged the questionnaires to observations taking place after the questionnaire was completed, and a 'mixed model' merged the questionnaires partly to retrospective observations and partly to prospective observations. A 'shortened mixed model' finally matched the questionnaires to only the nearest five minutes of observations before and after the questionnaire. The best model fit was achieved by both mixed models with a slightly better fit for the shortened model, indicating that emotional state, perceived difficulty, and current task enjoyment relate to on-task behaviour both in the minutes before and immediately after the self-report is completed. As the observation periods in this study were shorter than in the MSc study, it was decided to retain the data from the full observation periods.

Thus, in this study the observations in the first 7.5 minutes of the lesson were merged with the first questionnaire, the middle five to fifteen minutes were merged with the second questionnaire, and the final 7.5 minutes with the third questionnaire. As lessons varied slightly in length, the first and last 7.5 minutes were always matched with the first and third questionnaire respectively, and the remaining middle section with the second questionnaires. Only observed lessons of at least 40

observations were included in analyses, meaning the middle section was never shorter than five minutes (ten observations). The maximum observation period was 30 minutes, leading to a maximum duration of the middle section of fifteen minutes (30 observations).

### 3.10.3 Analytic Strategies

#### Multilevel Linear Models

Multilevel linear modelling was used to analyse the data to account for its hierarchical structure. Due to the DV being binary, logistic models with a logit link were used (Field, 2013). In the current dataset, observations were nested within weeks, weeks in participants, and participants in classrooms (Figure 13). Weeks, participants, and classrooms are contextual variables; they are common to a group of observations and make them more alike. Thus, contextual variables lead to a dependency in the data, and exclude the use of linear regression models for analysis (ibid.).

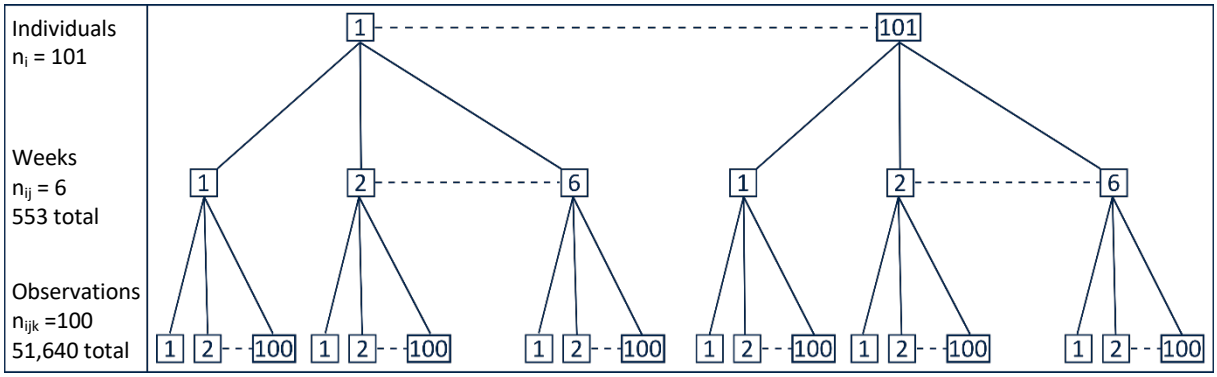



*Figure 13. Multilevel structure of observation data.*

Nested data also brings the potential for sphericity, as timepoints which are close in time (within the same day) are more likely to be similar than those far apart (from week to week) (Tabachnick and Fidell, 2013). The same can be said of data nested in persons and persons in groups; the data from repeated observations of one participant will be more similar to each other, than to the observations of another participant. Additionally, the data was collected in eight different classrooms, making observations of children in the same classroom more likely to be similar to each other than to those

in other classrooms. If nested data is analysed using non-MLM statistics, type I error rates can be greatly increased and the significance of intervention effects overstated (ibid.).

To assess the proportion of variance in the outcome variable related to each level of the model, the ICC is calculated (Field, 2013) as a measure of within-group similarity. A small ICC indicates that there is little similarity within each group and thus little variation between groups, whereas a large ICC indicates a great deal of similarity within groups and great variation between groups. Therefore, the larger the ICC is, the more important it is to include the particular level in the ICC. In this way, the ICC can help to establish if there is a need for MLM, and if additional levels should be added to the model (Tabachnick and Fidell, 2013).

To calculate the ICC, an intercepts-only model (or null-model) without predictor variables is run. See Equation 6 for the three-level null model.

Subscripts: i = timepoint, j = week, k = participants

Level 1: $Y_{ijk} = \beta_{0jk} + e_{ijk}$

$Y_{ijk}$ = on-task behaviour at time-point i in week j for participant k

$\beta_{0jk}$ = intercept: mean behaviour in week j for participant k

$e_{ijk}$ = variance: time-point specific residual variance at time-point i in week j for participant k

Level 2: $\beta_{0jk} = \delta_{00k} + u_{0jk}$

$\beta_{0jk}$ = mean behaviour in week j for participant k

$\delta_{00k}$ = mean behaviour for participant k

$u_{0jk}$ = variance: difference between participant k's behaviour in week j and their average behaviour across all six weeks

Level 3: $\delta_{00k} = \gamma_{00} + v_{00k}$

$\delta_{00k}$ = mean on-task behaviour for participant k

$\gamma_{00}$ = grand mean (sample mean behaviour)

$v_{00k}$ = variance: person-specific deviance from the grand mean

When all of these are put together into one equation, it becomes:

Combined: $Y_{ijk} = b_0 + v_{0k} + u_{0jk} + e_{0ijk}$

*Equation 6. Null Model.*

To find the ICC for each level, the variance at the targeted level is divided by the sum of the variances at all levels (Tabachnick and Fidell, 2013). The equations to calculate the ICC for the three-level null-model can be found in Equation 7, whereby $\rho$ indicates the variance at a certain level. In this study, the levels were timepoints, weeks, and participants.

$$ICC_{weeks} = \frac{\rho_{weeks}}{\rho_{timepoints} + \rho_{weeks} + \rho_{participants}} = \frac{\rho_{weeks}}{\left(\frac{\pi^2}{3}\right) + \rho_{weeks} + \rho_{participants}}$$

$$ICC_{participants} = \frac{\rho_{participants}}{\rho_{timepoints} + \rho_{weeks} + \rho_{participants}} = \frac{\rho_{participants}}{\left(\frac{\pi^2}{3}\right) + \rho_{weeks} + \rho_{participants}}$$

*Equation 7. Intra Class Correlation coefficient.*

As the DV is binary, variance at the lowest level in logistic models is not calculated, but approximated by using $\frac{\pi^2}{3}$ (Goldstein, Browne and Rasbash, 2002). This is then used in pseudo ICC calculations at each of the levels of the MLM. These pseudo ICC reflect the relative amounts of variance at each level, but, unlike in linear multilevel regression, do not represent a percentage of variance. The pseudo ICCs for the three-level null model were calculated, using the null model equation.

After calculating the pseudo ICC at each level, it was found that the amount of variance differed between levels, and MLM was confirmed as the appropriate analysis strategy. Random intercepts and slopes, as well as predictor variables were added at the various levels (see section 3.10.4 for the model building strategy). As the effect of each intervention condition was formed of a combination of parameters and cross-level interaction parameters, Wald tests were used in MLwiN to assess the significance of each intervention condition (Leckie and French, 2013).

## Structural Equation Models

Structural Equation Models and multilevel SEM (MSEM) were used to analyse the data from participants' self-reported emotions. SEM was selected for this analysis for two reason. First, SEM allows for analysis of both observed and unobserved (latent) variables (Byrne, 2009), which applies to this study because behaviour and PA were observed variables, and emotion states were latent variables. The second reason for using SEM is that it is the most widely applied method for analysing indirect effects (ibid.), and emotional state was hypothesised to have an indirect effect on on-task behaviour. An additional benefit of SEM is that it allows for simple visualisations of the specified models (ibid.). Observed variables are represented by square or rectangular boxes, latent variables by round or oval boxes, single-headed arrows denote regression paths, and double-headed arrows correlation or covariance (Geiser, 2013).

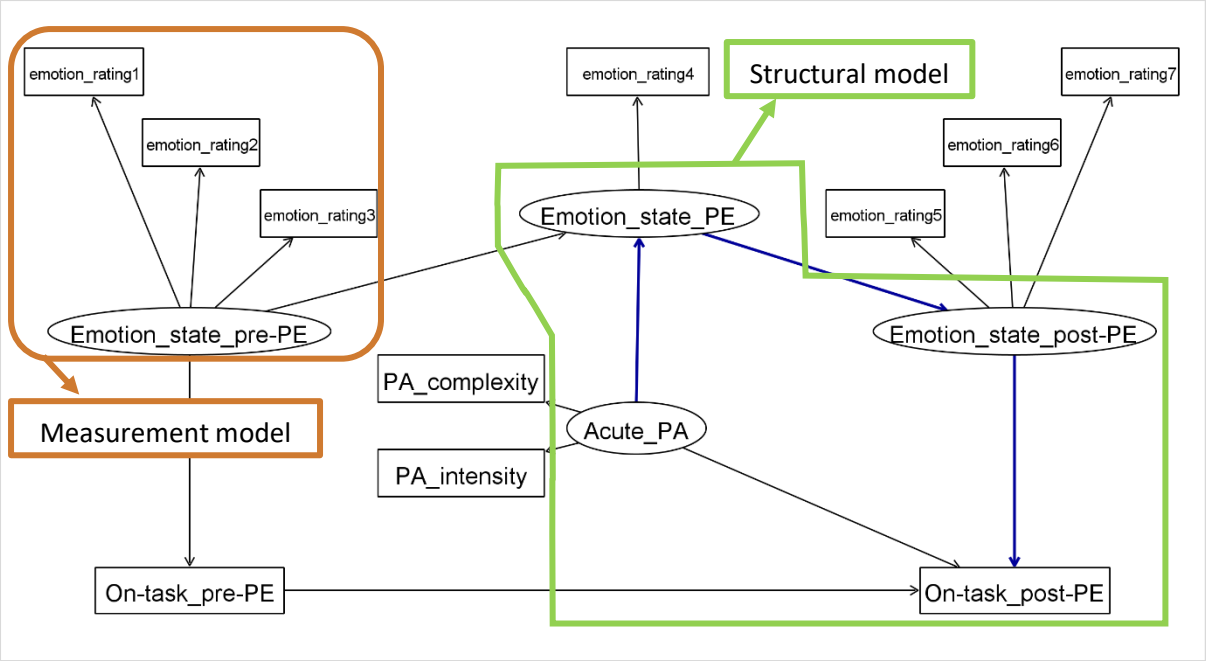



*Figure 14. The structural and measurement parts of SEM.*

To combine observed and latent variables, SEM consist of two parts; the structural model, and the measurement model (Byrne, 2009) (Figure 14). The structural part comprises the hypothesised relationships between latent (and/or observed) factors, whilst the measurement part of the model relates the latent factors to the variables they are measured by (Tabachnick and Fidell, 2013). The

measurement part of the model is decided by factor analysis; a process whereby observed variables which are correlated with each other but mostly uncorrelated with other variables in the dataset, are grouped together into 'factors' (ibid.). These factors are assumed to represent underlying processes or states that have created the correlations in the variables (ibid.).

Two types of factor extraction exist; exploratory and confirmatory factor analysis (EFA and CFA respectively). EFA is typically used for newly developed data collection instruments, or if the relationships between variables is unknown or uncertain, to assess how and to what extent the items in the instrument are related to the underlying latent factors which are intended to be measured (Byrne, 2009). CFA is used when a certain structure in the data is expected or known, based on theory or previous experience (ibid.). The structure of the measurement model is specified a priori in CFA, and tested for 'goodness of fit' (ibid.).

The second reason for the use of SEM was that it is ideally suited to the analysis of indirect effects. Indirect effects are also known as mediation effects, as the effects are mediated through another variable (Geiser, 2013). Mediator variables are dependent and independent variables at the same time (ibid.), as demonstrated in Figure 15. In the relationship between the IV and DV without mediation (Figure 15A.), the DV Y is predicted by IV X, via path c (Equation 8) (Preacher *et al.*, 2007). When a mediator (M) is introduced (Figure 15B.), X predicts M via path $a_1$, M predicts Y via path $b_1$, and X predicts Y via path c' (Equation 9) (ibid.). The mediation effect is equal to the difference between c and c', which in turn is equal to $a_1 * b_1$ (Equation 10) (ibid.).

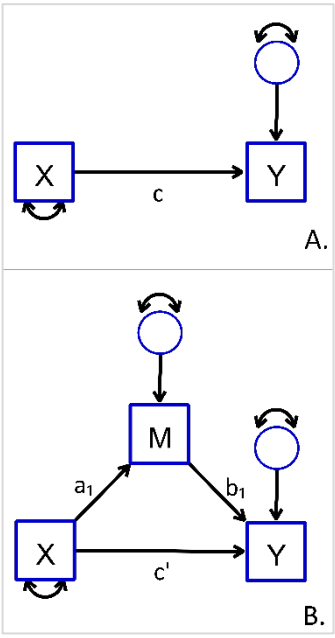



*Figure 15. A. Direct effect of X on Y. B. Effect of X on Y, mediated by M (Preacher, Rucker and Hayes,*

$$Y = c_0 + cX + residual$$

*Equation 8. Regression without mediator.*

$$Y = b_0 + c'X + b_1M + residual$$

$$M = a_0 + a_1X + residual$$

*Equation 9. Regression with mediator.*

$$a_1 * b_1 = c - c'$$

*Equation 10. Mediation effect.*

#### Analysis of Variance

Analysis of variance (ANOVA) is used to compare the mean scores between 3 or more groups (Field, 2013). The aim of an ANOVA is to assess if a model based on group means is more representative of the real data than a model based on the grand mean (ibid.). This is the case when the differences between group means are sufficiently large (ibid.).

ANOVA was selected for the analysis of the accelerometery data to identify if there was significant difference between the three intensity levels of the intervention lessons (low, medium, and high) in respect of the mean time spent at each PA intensity (light, moderate, and vigorous). Although ANOVA ignores the hierarchy in the data, in this analysis it was of interest how much movement occurred across the sample and relative to established cut-off points, rather than how much a certain participant moved relative to their own previous movement level, or any other participants'.

### 3.10.4 Application of Analytic Strategies

#### Analyses of Behaviour Observations

First, descriptive statistics of the observation data were obtained, and the mean proportion of on-task behaviour was compared against benchmarks from the existing literature. The data was then split into pre- and post-PE lessons, and descriptives were repeated. To investigate if differences in pre- and post-PE task behaviour were related to any of the collected control variables, observed behaviour was cross-tabulated with task type, task enjoyment, task difficulty, tiredness, attainment level, gender, year group, BMI category, meeting government guidelines for free-living MVPA, and frequency and duration of participation in organised sport. Correlations between the control variables and task

behaviour were analysed with Kendall's Tau B as all variables were binomial, categorical, or ordinal (Field, 2013). Significantly related variables were included in the MLM as control variables.

Three-level models were specified in MLwiN (Charlton *et al.*, 2019) using the long-format dataset. The probability of observations being on-task was modelled, using binomial models with a logit link. Equation 11 to Equation 17 demonstrates the step-wise building of these models. In all model equations the subscript 'i' indicates the observation level, 'j' the week level, and 'k' the participant level.

*Step 1*. The first predictor added into the model was whether the observation was taken pre- or post-PE, to see if on-task behaviour differed before and after PE lessons (Equation 11). This was added as a fixed parameter initially (PrePost), and random intercepts and slope were added. The intercept for on-task behaviour was expected the vary from participant to participant, and also from week to week, so it was added as a random effect at level 2 and 3 (Equation 12). As the change in on-task behaviour from before to after PE was expected to differ from week to week, it was added as a random variable at the week level (level 2, Equation 13). It was hypothesised that those participants with the lowest on-task behaviour before PE show a greater increase in on-task behaviour than those with higher on-task behaviour before PE, thus a negative covariance for intercept and slope was expected at level 2.

$$On-task_{ijk} = \beta_0 + \beta_1 PrePost_{ijk} + \nu_{0k} + \upsilon_{0jk}$$

*Equation 11. Model 1.1: Fixed effects.*

$$On-task_{ijk} = \beta_{0jk} + \beta_1 PrePost_{ijk}$$
$$\beta_{0jk} = \beta_0 + \nu_{0k} + \upsilon_{0jk}$$
$$intercept\ variance\ at\ level\ 3\text{: } [\nu_{0k}] \sim N(0, \Omega)\text{: } \Omega_\nu = [\sigma^2_{\nu 0}]$$
$$intercept\ variance\ at\ level\ 2\text{: } [\upsilon_{0jk}] \sim N(0, \Omega_\upsilon)\text{: } \Omega_\upsilon = [\sigma^2_{\upsilon 0}]$$

*Equation 12. Model 1.2: Random intercepts.*

$$On-task_{ijk} = \beta_{0jk} + \beta_{1j} PrePost_{ijk}$$
$$\beta_{0jk} = \beta_0 + \nu_{0k} + \upsilon_{0jk}$$
$$\beta_{1j} = \beta_1 + \upsilon_{1jk}$$

$$intercept\ variance\ at\ level\ 3\text{: } [\nu_{0k}] \sim N(0, \Omega)\text{: } \Omega_\nu = [\sigma^2_{\nu 0}]$$

$$\begin{bmatrix} intercept\ variance\ level\ 2 & \\ intercept - slope\ covariance & slope\ variance\ level\ 2 \end{bmatrix} = \begin{bmatrix} \upsilon_{0jk} \\ \upsilon_{1jk} \end{bmatrix} \sim N(0, \Omega_\upsilon)\text{: } \Omega_\upsilon = \begin{bmatrix} \sigma^2_{\upsilon 0} & \\ \sigma_{\upsilon 01} & \sigma^2_{\upsilon 1} \end{bmatrix}$$

*Equation 13. Model 1.3: Random intercepts and slope.*

Control variables were then added one by one, and the effect size of PrePost calculated after the addition of each control variable. The full final equation for model one can be found in Equation 14.

$$On - task_{ijk} = \beta_{0jk} + \beta_{1j} PrePost_{ijk} + \beta_{2-28} control\ variables$$

$$\beta_{0jk} = \beta_0 + \nu_{0k} + \upsilon_{0jk}$$

$$\beta_{1j} = \beta_1 + \upsilon_{1jk}$$

$$[\nu_{0k}] \sim N(0, \Omega)\text{: } \Omega_\nu = [\sigma^2_{\nu 0}]$$

$$\begin{bmatrix} \upsilon_{0jk} \\ \upsilon_{1jk} \end{bmatrix} \sim N(0, \Omega_\upsilon)\text{: } \Omega_\upsilon = \begin{bmatrix} \sigma^2_{\upsilon 0} & \\ \sigma_{\upsilon 01} & \sigma^2_{\upsilon 1} \end{bmatrix}$$

*Equation 14. Model 1.4: PrePost with control variables.*

*Step 2*. In step 2, hypothesis 1 was tested. It was hypothesised that there is a main effect of PE lesson complexity on post-PE on-task behaviour. The model was expanded with a level 2 predictor variable (complexity), and a cross-level interaction of complexity*PrePost (Equation 15). For clarity, individual control variables are omitted from the equation.

$$On - task_{ijk} = \beta_{0jk} + u_{1jk} PrePost_{ijk} + \beta_2 Complexity_{jk} + \beta_3 PrePost * Complexity_{ijk} + \beta_{4-30} control\ variables + \nu_{0k} + \upsilon_{0jk}$$

*Equation 15. Model 2 (RQ1).*

*Step 3*. Hypothesis 2, that there is a main effect of PE lesson intensity on subsequent on-task behaviour, was tested in step 3. The level 2 predictor complexity was replaced with two dummy-coded variables for medium intensity and high intensity, and the appropriate cross-level interactions (Equation 16).

$$On - task_{ijk} = \beta_{0jk} * cons + u_{1jk} PrePost_{ijk} + \beta_2 Intensity_{medium\,jk} + \beta_3 Intensity_{high\,jk} + \beta_4 PrePost * Intensity_{medium\,ijk} + \beta_5 PrePost * Intensity_{high\,ijk} + \beta_{6-32} control\ variables + \nu_{0k} + \upsilon_{0jk}$$

*Equation 16. Model 3 (RQ2).*

*Step 4*. Finally, hypothesis 3, the predicted interaction effect of PE lesson intensity and complexity on post-PE on-task behaviour, was tested. the interaction of intensity and complexity was modelled, along with the cross-level interactions with PrePost (Equation 17).

$$
\begin{aligned}
On-task_{ijk} = {} & \beta_{0jk} * cons + u_{1jk} PrePost_{ijk} + \beta_2 Intensity_{medium\,jk} + \beta_3 Intensity_{high\,jk} \\
& + \beta_4 PrePost * Intensity_{medium_{ijk}} + \beta_5 PrePost * Intensity_{high_{ijk}} \\
& + \beta_6 Complexity_{jk} + \beta_7 PrePost * Complexity_{ijk} \\
& + \beta_8 Intensity_{medium} * Complexity_{jk} + \beta_9 Intensity_{high} * Complexity_{jk} \\
& + \beta_{10} PrePost * Intensity_{medium} * Complexity_{ijk} + \beta_{11} PrePost \\
& * Intensity_{high} * Complexity_{ijk} + \beta_{12-38} control\ variables + \nu_{0k} + \upsilon_{0jk}
\end{aligned}
$$

*Equation 17. Model 4 (RQ3).*

At each of the steps 1-4, the analyses were repeated with stepwise addition of eight control variables: teacher-reported attainment, year group, task type, gender, self-reported task enjoyment, BMI category, self-reported weekly duration of sports participation relative to sample mean, and self-reported task difficulty.

$$
\begin{aligned}
On-task_{ijk} = {} & \beta_{0jk} * cons + u_{1jk} PrePost_{ijk} + \beta_2 Intensity_{medium\,jk} + \beta_3 Intensity_{high\,jk} \\
& + \beta_4 PrePost * Intensity_{medium_{ijk}} + \beta_5 PrePost * Intensity_{high_{ijk}} \\
& + \beta_6 Complexity_{jk} + \beta_7 PrePost * Complexity_{ijk} \\
& + \beta_8 Intensity_{medium} * Complexity_{jk} + \beta_9 Intensity_{high} * Complexity_{jk} \\
& + \beta_{10} PrePost * Intensity_{medium} * Complexity_{ijk} + \beta_{11} PrePost \\
& * Intensity_{high} * Complexity_{ijk} + \beta_{12} Gender + \beta_{13} PrePost * Gender \\
& + \beta_{14} Intensity_{medium} * Gender + \beta_{15} Intensity_{high} * Gender + \beta_{16} Complexity \\
& * Gender + \beta_{17} Intensity_{medium} * Complexity * Gender + \beta_{18} Intensity_{high} \\
& * Complexity * Gender + \beta_{19} PrePost * Intensity_{medium} * Gender \\
& + \beta_{20} PrePost * Intensity_{high} * Gender + \beta_{21} PrePost * Complexity * Gender \\
& + \beta_{22} PrePost * Intensity_{medium} * Complexity * Gender \\
& + \beta_{23} PrePost * Intensity_{high} * Complexity * Gender \\
& + \beta_{24-46} control\ variables + \nu_{0k} + \upsilon_{0jk}
\end{aligned}
$$

*Equation 18. Model with interaction effects for gender.*

Literature suggests that those with moderate fitness levels (Chang *et al.*, 2014) or higher fitness levels (Jäger *et al.*, 2015), those with higher academic achievement (ibid.), and those with lower trait

on-task behaviour (Ma *et al.*, 2014; Mahar *et al.*, 2006) are most likely to benefit from acute PA interventions aimed at increasing cognitive function or time on task. To test hypothesis 4, and investigate if the effects of intensity and complexity varied between participants in this sample, interaction effects of intervention conditions by attainment, mean pre-PE probability of being on-task, and average weekly duration of sports participation were added to the model. Finally, although gender differences do not feature strongly in the literature, interactions with gender were examined. The same control variables were used in these analyses, but, as in some of these analyses one control variable became a predictor variable, the total number of control variables was seven or eight. Equation 18 gives an example of the additional interaction effect specification for gender. In this example, gender was added to interact with each of the other predictors, in all possible combinations, leading to the addition of twelve parameters to the model.

### Analyses of Self-Reported emotions

This study utilised a newly developed tool for the collection of emotion state data. As the hypothesized factor structure had not been previously tested, EFA was conducted in Mplus8 (Muthen and Muthen, 2017). The eight self-reported emotion items were included, and oblique rotation applied as the emotion factors were expected to correlate (Field, 2013). The hypothesised four-factor structure (Figure 16) was not indicated by the data, and a two-factor structure based on six emotions was found instead. EFA did not allow for modelling of the multilevel structure of the data, and multilevel confirmatory factor analysis (MCFA) was conducted to verify the obtained factor structure in the hierarchical data structure.

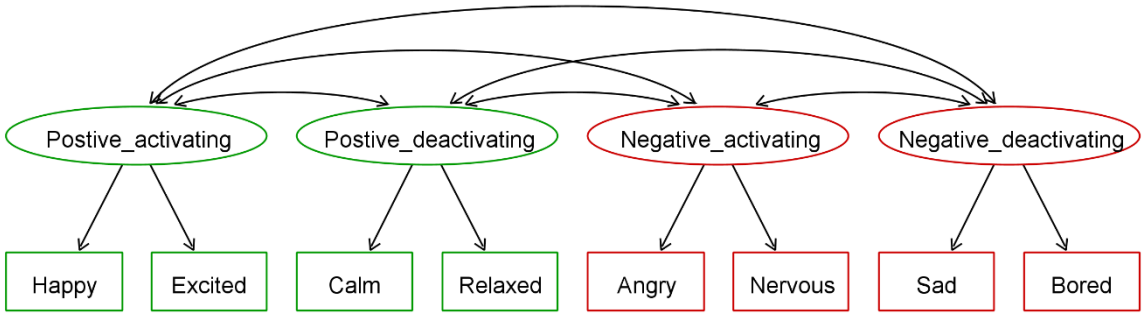



*Figure 16. Hypothesised factor structure.*

First, the structure discovered in EFA was replicated in a three-level model. Next, using the wide dataset with the answers of the seven repeated questionnaires on each line of data for a single week, invariance testing was performed on a two-level model. Then, correlations between questionnaires at various points in the day were examined, and an autoregressive model was run to check if emotion scores at a given point could be predicted by previous emotion scores. Indirect paths were included from the pre-PE to the post-PE lesson via the PE lesson to see if emotions after PE were mediated by emotions during the PE lesson.

For further models, emotion scores were aggregated at the questionnaire-level to reduce the number of parameters in the model and allow for the addition of predictor variables, as well as on-task behaviour as the outcome variable. This structure was then applied to the regression model, and indirect paths from pre-PE to post-PE via the PE lesson were added. This model examined the predictive value of pre-PE emotions and PE-lesson emotions on post-PE emotions. Next, cross-lagged paths from positive pre-PE to negative PE, and positive post-PE, as well as negative pre-PE to positive PE, and negative post-PE were specified (Figure 17). The cross-lags sought to establish if participants reporting greater positivity or negativity before PE were generally more positive or negative, or if those reports were related differently to the context (classroom or PE).

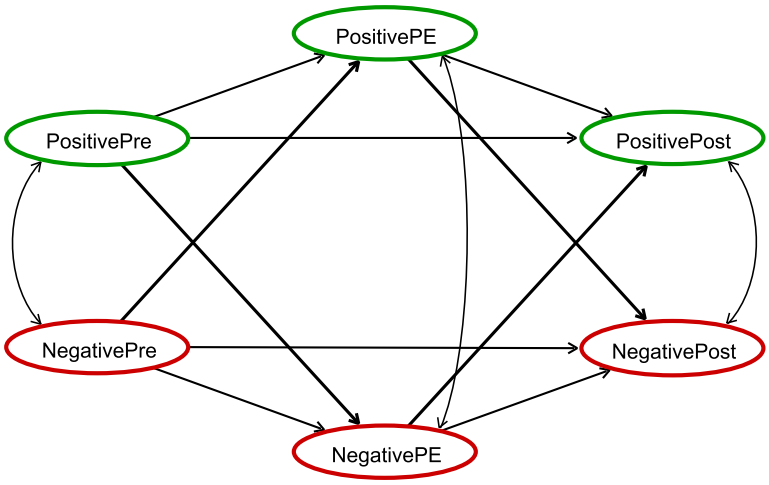



*Figure 17. Simplified model with cross-lags.*

Predictor variables from the PE lessons (PE lesson conditions and self-reported PE enjoyment) were added to the model, along with indirect paths from PE lesson predictors to post-PE emotion states via PE lesson emotion states, to see if PE lesson content or subjective enjoyment of the PE lesson could predict emotional state in the classroom after PE. Finally, aggregated on-task behaviour variables were added to the model. A pre-PE proportion on-task was included as well as a post-PE proportion on-task, with post-PE on-task predicted by pre-PE on-task, PE lesson condition, PE lesson enjoyment, and positive/negative emotions during the post-PE lesson. Emotions were predicted by PE lesson condition and PE enjoyment. Therefore, emotion state was a DV as well as an IV, and double mediation paths were specified from acute PA (PE lesson content) and PE enjoyment to on-task behaviour post-PE via emotion states (Figure 18 and Figure 19).

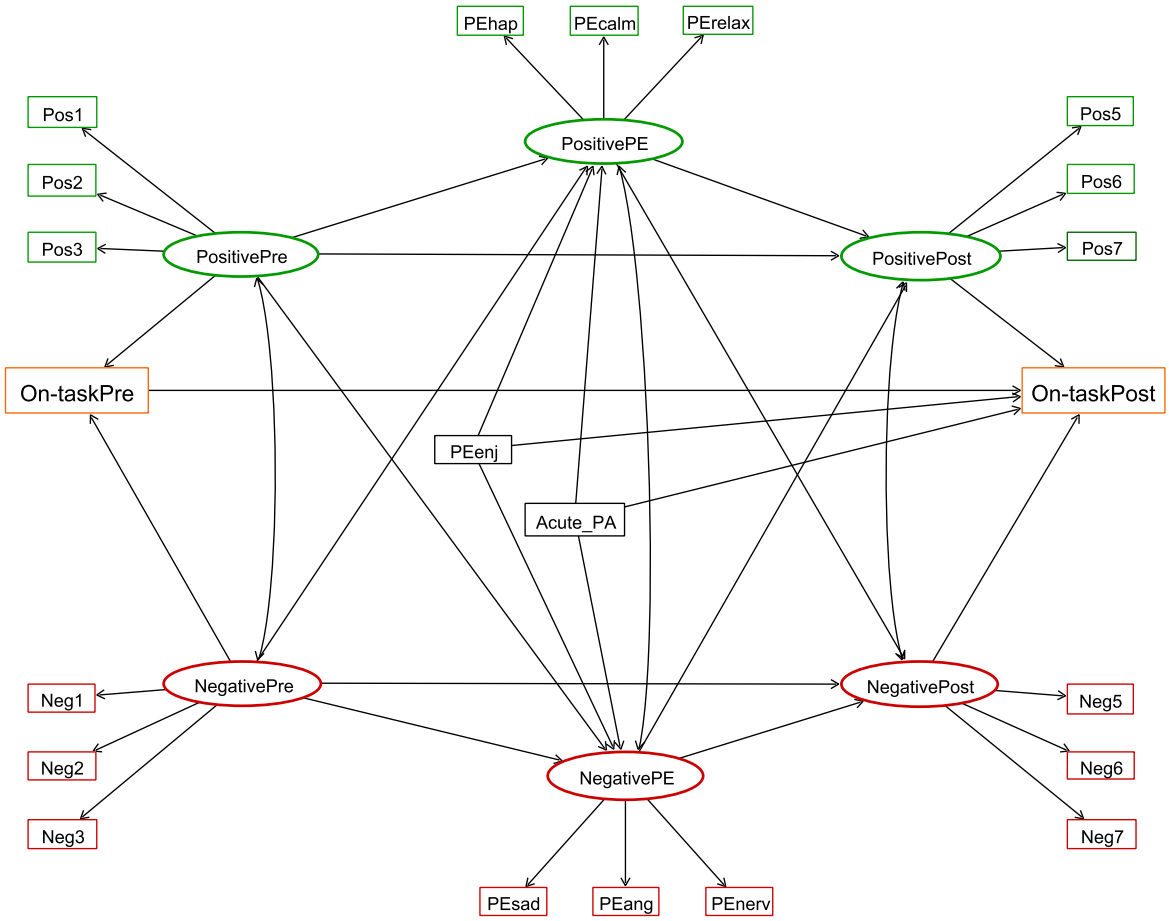



*Figure 18. Simplified full SEM model.*

The mediation paths in the model are highlighted in Figure 19. Equation 19 – Equation 21 demonstrate how the mediation effect of emotion states in the relationship between acute PA and post-PE on-task behaviour is calculated, following notation from Hayes and Preacher (2014) and Hayes (2018). Parameters with subscript '0' are intercepts, and emotion state can be either positive or negative.

$$PostPE\ on-task\ behaviour = constant + c_1 * Acute\ PA + residual$$

*Equation 19.Direct effect of acute PA on post-PE on-task behaviour.*

$$PostPE\ on-task\ behaviour = b_0 + b_1 Emotion\ state\ postPE + c_1' * Acute\ PA + residual$$

$$Emotion\ state\ postPE = d_0 + d_{21} Emotion\ state\ PE + residual$$

$$Emotion\ state\ PE = a_0 + a_1 Acte\ PA + residual$$

*Equation 20. Mediated effect of acute PA on post-PE on-task behaviour.*

$$a_1 * d_{21} * b_1 = c_1 - c_1'$$

*Equation 21. Mediated effect of acute PA on on-task behaviour via emotion state.*

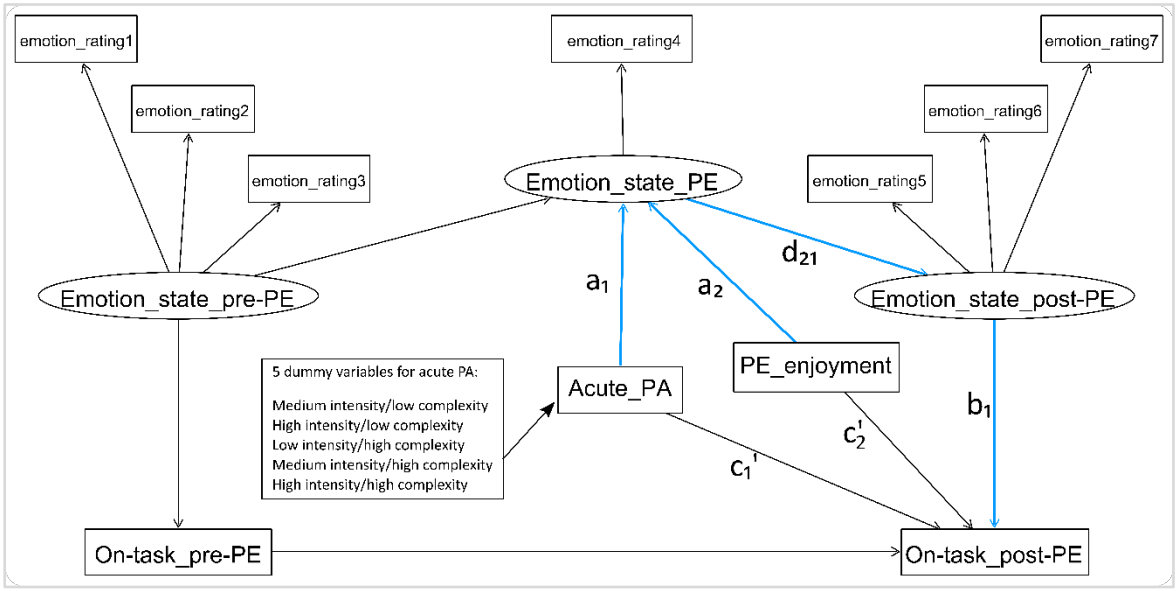



*Figure 19. Hypothesised mediation effect of emotional state on behavioural engagement*

Likewise, mediation of the relationship between PE enjoyment and on-task behaviour via emotion states is calculated according to Equation 22-Equation 24. Again, parameters with subscript '0' are intercepts, and emotion state can be either positive or negative.

$$PostPE\ on-task\ behaviour = constant + c_2 * PE\ enjoyment + residual$$

*Equation 22. Direct effect of PE enjoyment on post-PE on-task behaviour.*

$$PostPE\ on-task\ behaviour$$

$$= b_0 + b_1 Emotion\ state\ postPE + c_2' * PE\ enjoyment + residual$$

$$Emotion\ state\ postPE = d_0 + d_{21} Emotion\ state\ PE + residual$$

$$Emotion\ state\ PE = a_0 + a_2 PE\ enjoyment + residual$$

*Equation 23. Mediated effect of PE enjoyment on post-PE on-task behaviour.*

$$a_2 * d_{21} * b_1 = c_2 - c_2'$$

*Equation 24. Mediated effect of PE enjoyment on on-task behaviour via emotion state.*

### Analyses of Accelerometery Data and Activity Habits

A dataset was constructed with up to six rows per participant; each row representing a 24h recording of activity data for each week of the intervention they participated in. Variables included the number of minutes they spent at sedentary, LPA, MPA, and VPA, as well as their mean ENMO and minutes at MVPA for each section of the 24h recording: end of school until 8pm, 8pm until midnight, midnight until 7am, 7am until the start of school, start of school until lunch, lunch, pre-PE afternoon lesson, PE lesson, and post-PE afternoon lesson. Variables were added for the intended intensity (0 = low, 1 = medium, 2 = high) and complexity (0 = low, 1 = high) of the PE lesson.

*Table 4. Overview of physical activity data analysis strategy.*

| Dataset used | Outcome variable | Analysis performed |
|---|---|---|
| PE lesson accelerometery | MVPA duration | One-way ANOVA of MVPA (intensity) |
| | % MVPA of total lesson duration | Factorial 2x3 ANOVA (complexity x intensity) |
| | % MPA of total lesson duration | Factorial 2x3 ANOVA (complexity x intensity) |
| | % VPA of total lesson duration | Factorial 2x3 ANOVA (complexity x intensity) |
| | Mean ENMO | Factorial 2x3 ANOVA (complexity x intensity) |
| 24h accelerometery | MVPA duration | Comparison to government guidelines of 60min/day & create the binary control variable 'Active Child' based on meeting government guidelines on at least 75% of recorded days |
| PA habits questionnaire | Weekly duration of organised PA participation | Create the binary control variable 'Sports participation duration' |
| | Weekly frequency of organised PA participation | Create the binary control variable 'Sports participation frequency' |

*Intervention Fidelity*. As an intervention fidelity check, a one-way ANOVA was carried out on the PE lesson data. Lessons intended to be low, medium, and high intensity were compared against each other for duration at MVPA to establish if the intensity participants exercised at were truly reflective of the intended intensity. Pairwise comparisons were performed for each lesson intensity pair, using Gabriel's and Bonferroni's post-hoc tests. These were selected on the basis that they are reliable when

group sizes are unequal across groups, they control for type I error without too great a loss of power, and they are suitable for relatively small sample sizes and a low number of groups (Field, 2013).

Minutes at MVPA is the most commonly used metric for reporting objectively measured activity in PA related research, thus the initial intervention fidelity check was based on minutes at MVPA. However, lesson duration varied between intervention conditions, and it was decided to change the metric to the percentage of lesson time spent at MVPA. Next, a factorial 2 (complexity) by 3 (intensity) ANOVA was carried out to identify separately the effect of intensity and complexity, as well as the interaction effect and determine the variation in lesson intensity across all intervention conditions. Again, post-hoc analyses were conducted and the percentage of MVPA did not vary between all lesson intensities.

To further investigate the nature of PE lesson intensity, MVPA was split into MPA and VPA and 3-by-2 ANOVAs were repeated. It was discovered that conditions fluctuated in the ratio of MPA and VPA which made up the MVPA duration. Lessons which looked to be the same intensity based on the percentage of time spent at MVPA were found to differ in their ratio of MPA to VPA. On this basis, to make a better assessment of the relative overall intensity of the PE lesson, mean ENMO was considered to be a better metric.

The raw mean ENMO metric may not be reflective of the lesson intensity if compared to the cut points for LPA, MPA, and VPA due to the short burst nature of children's PA. And, especially in the case of higher intensity lessons, the inclusion of lesson elements such as the warm-up and plenary would reduce the mean ENMO metric. Yet, it may be a more appropriate metric to assess the cumulative intensity of a PE lesson, as it gives one score which includes the balance of light, moderate, and vigorous bursts over the entire time period. Furthermore, as it is a mean score it automatically adjusts for the duration of the measurement period. In this way, it allows for assessment of overall PA intensity which would be complicated to disentangle if PA intensities were looked at in isolation, and gives a more nuanced view compared to the consideration of only MVPA.

Specific to this study, the intensity of the lessons was purposely manipulated by varying how quickly activity bursts succeeded one another. For example, by decreasing or increasing the time participants had to wait for their turn the lesson was made more or less intense. The mean ENMO from a whole lesson can therefore give a good indication of how intense overall a lesson has been. Hence, a 3-by-2 ANOVA of mean ENMO carried out.

*Activity Habits.* Additionally, the total time at MVPA accumulated in each 24h recording was calculated, excluding the PE lesson. For each day, this was then rated as 'achieving government guidelines for MVPA' or 'not achieving government guidelines for MVPA'. A binary score was produced on the basis of the daily rating relative to government guidelines. If a participant achieved the government guideline of at least 60 minutes of MVPA on at least 75% of recorded days, they were rated an 'active child'. This binary score was used as a control variable for further analyses of behaviour, based on Coe *et al.* (2006) who found that children achieving 'healthy people' PA guidelines outperformed their less active peers on all tests of academic achievement.

The data from the trait PA questionnaire was checked for outliers and the number and all minutes of sports participation were added up across the week to create a 'frequency of sports participation' and a 'duration of sports participation' variable. Additionally, each participant was compared to the sample mean for their frequency and duration of participation, and two more dichotomous variables were created with a value of '0' reflecting below the sample mean participation, and a value of '1' reflecting participation at or above the sample mean.

## 3.11 Concluding remarks

This chapter has described in detail the methods used for data collection, and the strategies employed in analysing this data. Moreover, it has provided a rationale for the selection of these specific methods and analyses, as appropriate to the field of PA and cognition research. The results of

the analyses are presented in chapter 4, and will be discussed in relation to the current body of literature in chapter 5.

# 4 Results

This chapter provides descriptive details of the sample and the variables collected for each analysis. In section 4.1, sample descriptives are given for the full sample and the subsample included in analyses. Section 4.2 presents the results of the intervention fidelity check. Next, in section 4.3, descriptives of the variables included in the MLM are provided, and research questions one to four are answered. Finally, in section 4.4, descriptives for the SEM and the answer to research question five are given.

## 4.1 Sample

### 4.1.1 Sample Descriptives

The total sample consisted of 108 participants, of whom 101 were observed during lessons, and teachers reported achievement levels for 76 participants, rating their attainment compared to Age Related Expectation (ARE). Full sample descriptives for all three sample sizes can be found in Table 5. There were no significant differences between the three samples in terms of age, gender, handedness, and BMI categories.

In the sample of 76 participants with attainment scores, age ranged from 7.9 years to 10.4 years (M=9.2 years, SD=0.7 years). Forty-six participants were female (60.5%) and 30 were male. The majority (88.2%) were right-handed. Thirty-four participants (44.7%) were rated ‘at ARE’. In this sample, 56.6% had a healthy or low-to-healthy BMI, with the remaining participants being above healthy, overweight, and obese. A quarter of the observed sample with attainment scores was overweight or obese. Distribution was not equal across the four schools, with more children in school 3 in the ‘above healthy’, ‘overweight’, and ‘obese’ categories.

*Table 5. Sample descriptives*

| | | Total sample (N=108) | Observed sample (N=101) | Observed sample with attainment (N=76) | Full vs. observed sample | Full vs. sample w/ attainment |
|---|---|---|---|---|---|---|
| Age | M (SD) | 9.3 (0.6) | 9.3 (0.6) | 9.2 (0.7) | t=0.044 | t=1.82 |
| | Range | 7.9 – 10.5 | 7.9 – 10.4 | 7.9 – 10.4 | df=106 | df=72.07 |
| | | | | | p>0.05 | p>0.05 |
| Gender | | | | | $\chi^2$=0.004 | $\chi^2$=0.004 |
| | Female | 65 (62.2%) | 60 (59.4%) | 46 (60.5%) | df=1 | df=1 |
| | Male | 43 (38.8%) | 41 (40.6%) | 30 (39.5%) | p>0.05 | p>0.05 |
| Handedness | | | | | $\chi^2$=0.062 | $\chi^2$=0.062 |
| | Right-handed | 95 (88.0%) | 89 (88.1%) | 67 (88.2%) | df=1 | df=1 |
| | Left-handed | 13 (12.0%) | 12 (11.9%) | 9 (11.8%) | p>0.05 | p>0.05 |
| Attainment | | | | | N/A | N/A |
| | Well below ARE | | | 5 (6.6%) | | |
| | Below ARE | | | 7 (9.2%) | | |
| | At ARE | | | 34 (44.7%) | | |
| | Above ARE | | | 20 (26.3%) | | |
| | Well above ARE | | | 10 (13.2%) | | |
| BMI category | | | | | $\chi^2$=6.436 | $\chi^2$=6.436 |
| Percentiles: | | | | | df=5 | df=5 |
| <2 | Low BMI | 1( 0.9%) | 0 (0.0%) | 0 (0.0%) | p>0.05 | p>0.05 |
| 2-25 | Low to healthy | 21 (19.4%) | 20 (19.8%) | 18 (23.7%) | | |
| 25-75 | Healthy | 42 (38.9%) | 39 (38.6%) | 25 (32.9%) | | |
| 75-91 | Above healthy | 17 (15.7%) | 17 (16.8%) | 14 (18.4%) | | |
| 91-98 | Overweight | 10 ( 9.3%) | 9 (8.9%) | 8 (10.5%) | | |
| >98 | Obese | 17 (15.7%) | 16 (15.8%) | 11 (14.5%) | | |

## 4.2 Intervention Fidelity

In this section, results from accelerometery are presented. There were two purposes to these analyses; to perform an intervention fidelity check, and to provide control variables for the multilevel analyses of on-task behaviour. For the fidelity check, accelerometery data from PE lessons were analysed. Section 4.2.1 sets out the descriptives for the PE accelerometery data, and results from the

fidelity check can be found in section 0. Accelerometery data was also recorded in order to control for participants' habitual PA intensity and duration. Data from the full 24 hours recorded was analysed to determine how active participants were relative to government guidelines for MVPA. The control variable analyses in relation to on-task behaviour are described in more detail in section 4.3.

### 4.2.1 Physical Education Lesson Accelerometery Descriptives

*Table 6. Duration at intensity by intervention lesson.*

| | | % MPA | | % VPA | | % MVPA | |
|---|---|---|---|---|---|---|---|
| Intensity | N | M | Std D | M | Std D | M | Std D |
| Low | 156 | 9.42 | 4.39 | 2.52 | 2.22 | 11.93 | 6.24 |
| Medium | 148 | 9.91 | 3.18 | 4.53 | 2.59 | 14.44 | 5.29 |
| High | 155 | 11.70 | 4.26 | 7.73 | 3.44 | 19.43 | 5.59 |
| Complexity | | | | | | | |
| Low | 234 | 7.85 | 2.71 | 4.36 | 3.65 | 12.22 | 5.46 |
| High | 225 | 12.94 | 3.68 | 5.51 | 3.31 | 18.45 | 6.00 |
| Intensity-by-Complexity | | | | | | | |
| Low intensity, low complexity | 77 | 6.71 | 2.21 | 1.50 | 1.70 | 8.21 | 3.44 |
| Medium intensity, low complexity | 76 | 8.04 | 2.69 | 3.21 | 1.91 | 11.25 | 4.18 |
| High intensity, low complexity | 81 | 8.76 | 2.81 | 8.17 | 2.98 | 16.94 | 4.51 |
| Low intensity, high complexity | 79 | 12.06 | 4.38 | 3.50 | 2.23 | 15.56 | 6.23 |
| Medium intensity, high complexity | 72 | 11.87 | 2.38 | 5.93 | 2.49 | 17.80 | 4.14 |
| High intensity, high complexity | 74 | 14.91 | 3.10 | 7.24 | 3.85 | 22.15 | 5.40 |

Note: PA thresholds from Sanders (personal communcation, 27 March 2019) and Fairclough *et al.* (2016). LPA: 56.3, MPA: 191.6, VPA: 695.8, MVPA: 191.6.

Mean valid duration of PE lessons, after recordings shorter than 30 minutes were excluded, was 45.5 minutes ($N_{PElessons}$=459, SD=5.2min, range=30–54.6). The percentage of lesson duration spent at

MPA, VPA and MVPA intensity, displayed by PE lesson intensity, complexity, and intervention condition, can be found in Table 6.

### 4.2.2 Intervention Fidelity: Accelerometery

Fluctuation in lesson duration was examined using factorial ANOVA and found to vary by intensity ($F_{(2,456)}$=10.17, p<0.001), complexity ($F_{(1, 457)}$=6.16, p<0.05), and intensity-by-complexity interaction ($F_{(2, 456)}$=5.62, p<0.01). For this reason, the fidelity check was conducted on MVPA duration measures as a percentage of lesson time.

#### Moderate-to-Vigorous Physical Activity by Lesson Intensity

First, a one-way ANOVA of MVPA intensities by PE lesson condition was carried out. The assumption of homogeneity of variance was not violated (Levene statistic$_{(2,456)}$=2.61, p=0.074). The percentage of time spent at MVPA was significantly different between lessons of different intensities ($F_{(2,456)}$=68.79, p<0.001), indicating a significant difference in duration of MPVA between the three PE lesson intensities. Post-hoc Bonferroni and Gabriel adjusted test scores were identical and are displayed in Table 7. Post-hoc tests revealed that all lessons differed significantly in MVPA duration.

*Table 7. Pairwise comparisons of MVPA duration during PE lessons*

| | Bonferroni & Gabriel post hoc contrasts | | |
|---|---|---|---|
| | Low-Medium | Low-High | Medium-High |
| % MVPA | -2.50 (0.66) *** | -7.49 (0.65) *** | -4.98 (0.66) *** |

Note: Mean difference (SE) p. Significance values *=p<0.05, **=p<0.01, ***=p<0.001

Factorial ANOVA: Intensity-by-Complexity

To investigate difference in lesson intensity across all six intervention conditions, a 3 (intensity) by 2 (complexity) ANOVA was carried out. This was to examine if the complexity level of the lessons may have influenced the MVPA measures for each lesson intensity. Results are shown in Table 8. Intensity

and complexity were significant for all PA measures, but the intensity-by-complexity interaction was not. The significant effect of complexity indicates that there may have been interference from the complexity level of the lessons on the intensity of PA recorded. The lack of a significant interaction effect of intensity-by-complexity suggests that some lesson conditions did not vary in the percentage of time spent at MVPA. This was further investigated.

*Table 8. 3-by-2 ANOVA results.*

| | df | F | p |
|---|---|---|---|
| Intensity | 2 | 103.96 | *** |
| Complexity | 1 | 206.34 | *** |
| Intensity-by-Complexity | 2 | 002.00 | |

Significance values *=p<0.05, **=p<0.01, ***=p<0.001

### Moderate-to-Vigorous Physical Activity by Lesson Condition

Figure 20 shows the sample means and 95% confidence interval of the percentage of lesson time spent at MVPA for each condition. It is clear that the lessons at medium intensity-high complexity do not differ much from the lessons at low intensity-high complexity, or from those at high intensity-low complexity. MVPA was split into its MPA and VPA components and the 3-by-2 ANOVA was repeated. The percentages of both MPA and VPA differed significantly by intensity (MPA: $F_{(2,456)}$=27.86, p<0.001, VPA: $F_{(2,456)}$=154.89, p<0.001), by complexity (MPA: $F_{(1,457)}$=326.63, p<0.001,

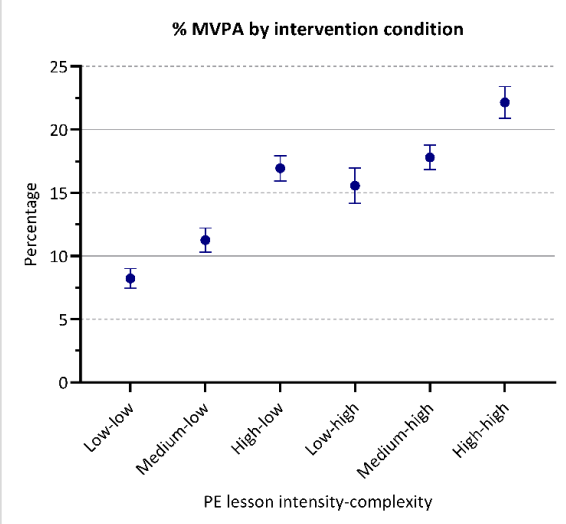



*Figure 20. Sample mean proportion and 95% confidence interval of PE lessons spent at MVPA.*

VPA: $F_{(1,457)}$=26.65, p<0.001), and by intensity-by-complexity (MPA: $F_{(2,456)}$=5.72, p<0.01, VPA: $F_{(2,456)}$=20.91, p<0.001).

The breakdown of MVPA by lesson condition is displayed in Figure 21. It shows that although low-high, medium-high, and high-low lessons did not differ greatly in their percentage duration at MVPA, this MVPA is made up of differing ratios of MPA and VPA. In particular, the proportion of VPA is greater in the high-low lesson compared to the medium-high condition, and greater in the medium-high compared to the low-high lesson.

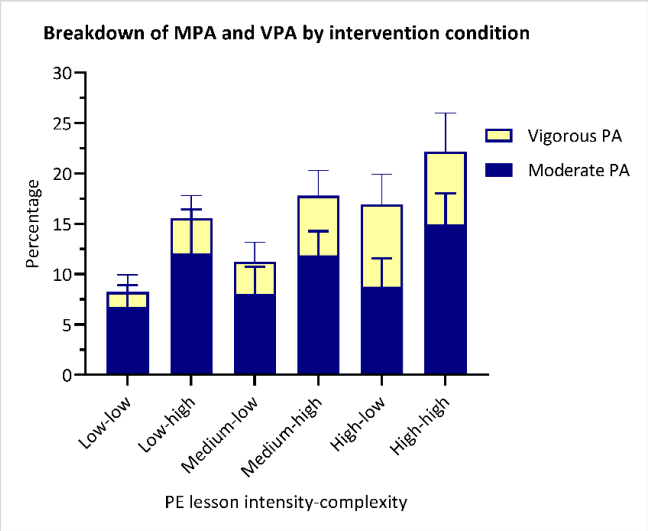



*Figure 21. Breakdown of moderate and vigorous PA by lesson condition.*

## Euclidean Norm Minus One

A 3(intensity)-by-2(complexity) ANOVA of mean ENMO during PE was carried out and results are reported in Table 9. Intensity, complexity, and the interaction of intensity and complexity have a significant effect on mean ENMO during PE lessons.

*Table 9. Factorial ANOVA of mean ENMO during PE.*

| | df | F | p |
|---|---|---|---|
| Intensity | 2 | 152.62 | *** |
| Complexity | 1 | 12.09 | ** |
| Intensity-by-Complexity | 2 | 45.01 | *** |

Significance values *=p<0.05, **=p<0.01, ***=p<0.001

Figure 22 shows the mean ENMO by PE lesson intensity. Mean ENMO for low intensity lessons was 147.06, which is in the LPA range (56.3-191.6). For medium intensity lessons mean ENMO was 216.87, and for high intensity lessons 316.27, both of which are in the MPA range (191.6-695.8). Although complexity did affect mean ENMO in PE lessons as well as the percentage of lesson duration at MVPA, Figure 23 shows that within the low complexity conditions there is a clear increase in mean ENMO between low, medium, and high intensity lessons, and the same pattern is observed for high complexity lessons.

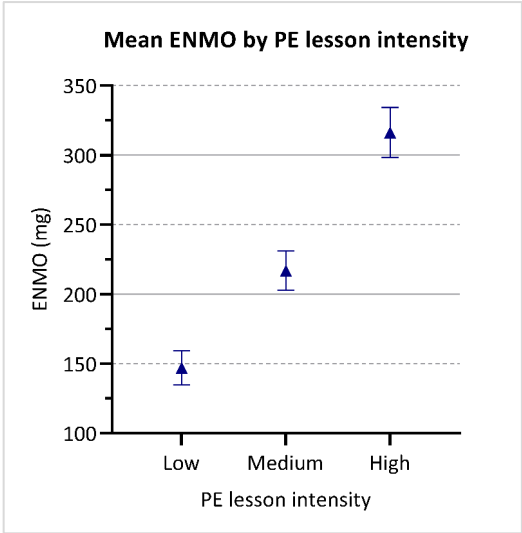



*Figure 22. Mean and 95% confidence interval of mean ENMO during PE lessons by lesson intensity.*

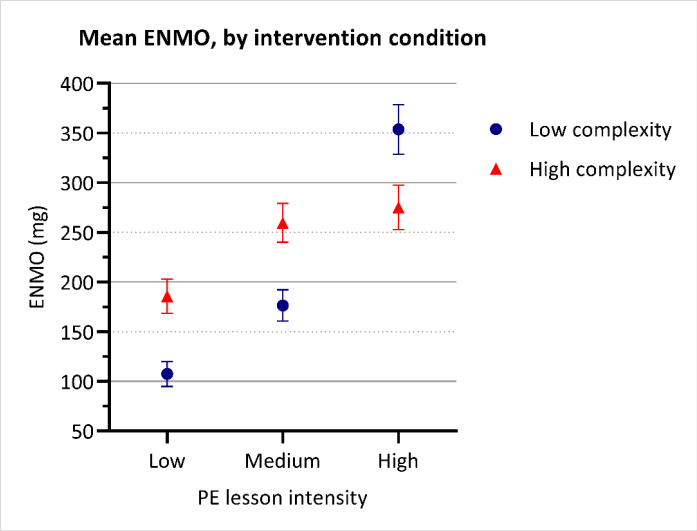



*Figure 23. Mean and 95% confidence interval of mean ENMO during PE lessons by PE condition.*

In conclusion, the low, medium, and high intensity lessons in the intervention differed significantly in the objectively measured intensity of PA. MVPA, MPA, VPA, and mean ENMO were all significantly higher in medium intensity than in low intensity lessons, and higher in high intensity lessons than in medium intensity lessons. The mean ENMO in low intensity lessons was in the LPA range, in medium and high intensity lessons it was in the MPA range. However, high intensity PE lessons had a

significantly higher mean ENMO than medium intensity lessons. There was some influence of lesson complexity on the recorded PA intensity, such that the intervention conditions all differed significantly in overall intensity.

## 4.3 Physical Activity and Behavioural Engagement

Descriptives of the variables included in the multilevel analyses and analytic results are specified in this section. Section 4.3.1 details the descriptives. First, descriptives are presented of the frequency and distribution of observations across the intervention conditions and classroom lessons. Next, behaviour descriptives are displayed at the sample level, and at the individual participant level. The occurrence of task type is compared between pre- and post-PE lessons, and control variable descriptives and correlations with on-task behaviour are shown. In section 4.3.2, results from the MLM are presented for each step of the model building, providing answers to research questions one to four.

### 4.3.1 Observations Descriptives

#### Frequency and distribution of observations

Table 10 sets out the distribution of observations across pre- and post-PE lessons, the six intervention conditions, and the three age groups observed for both the full observed sample and the observed sample with attainment scores. Only lessons where the participant had at least 40 valid observations (M = 50.6, SD = 3.3, range = 40-60) were included in analyses, giving a total of 51,640 observations for the full sample of 101 children, and 38,933 for the subsample of 76 participants with attainment scores.

There were significant differences between the full sample and sub-sample in how the observations were distributed across pre- and post-PE lessons, the conditions, and the year groups. However, as attainment was found to be an important covariate, it was retained in analyses and

descriptives are reported separately for this sub-sample. The majority of observations for the sub-sample were recorded in year 4 classes (51.3%), 16.1% of observations were of year 3 pupils, and 32.7% of year 5 pupils. Pre-PE observations accounted for 47.5% of the total, and the percentage of observations ranged between 16.3 and 17.5 across the intervention conditions.

*Table 10. Distribution of observations across intervention conditions and year groups*

| | Observed sample ($N_{observations}$=51,640) | Observed sample w/attainment score ($N_{observations}$=38,933) | Observed vs. w/ attainment t (p) |
|---|---|---|---|
| Pre-/post-PE | | | -11.61 (***) |
| Pre-PE | 48.9% | 47.5% | |
| Post-PE | 51.1% | 52.5% | |
| Intervention condition | | | -9.54 (***) |
| Low intensity, low complexity | 16.0% | 16.4% | |
| Medium intensity, low complexity | 15.7% | 16.3% | |
| High intensity, low complexity | 17.3% | 16.5% | |
| Low intensity, high complexity | 15.6% | 16.4% | |
| Medium intensity, high complexity | 17.8% | 17.5% | |
| High intensity, high complexity | 17.7% | 17.4% | |
| Year group | | | 60.79 (***) |
| Year 3 | 12.2% | 16.1% | |
| Year 4 | 50.5% | 51.3% | |
| Year 5 | 37.3% | 32.7% | |

Significance values *=p<0.05, **=p<0.01, ***=p<0.001

### Behaviour descriptives

Across all observations, the proportion of observed on-task behaviour was within the expected range based on existing literature, at 0.74 (SD=0.44) (Godwin and Fisher, 2011) and did not differ between the full sample and the subsample with attainment scores $t_{(51,638)}$=0.213, p=0.832. When observations were split between before and after PE, there was a difference between the full sample and the sub-sample. Descriptives for both samples are presented, and those for the sub-sample are

reported in more detail. Before PE lessons the overall proportion on-task was 0.72, and after PE it was 0.77 (Table 11).

Participants' individual mean task behaviour was calculated across all (up to twelve) observed lessons (Table 11). A large inter-individual range of on-task behaviour existed in the sample, with the least engaged participant only displaying on-task behaviours for 46% of their observed time, opposed to 91% for the most engaged participant. Active off-task behaviour was as high as 39% for the most off-task individual, with passive off-task behaviour ranging up to 29%. Mean other behaviour was low at 3%, but ranged up to 17%.

*Table 11. Behaviour descriptives whole sample and by ID.*

| | | | Observed sample (N=101) | | Observed sample with attainment score (N=76) | | Observed sample vs. w/attainment |
|---|---|---|---|---|---|---|---|
| | Behaviour | Lesson | M (SD) | Range | M (SD) | Range | t (p) |
| Sample aggregated | On-task | Overall | 0.74 (0.44) | | 0.74 (0.44) | | 0.21 |
| | | Before PE | 0.72 (0.45) | | 0.72 (0.45) | | 4.64 (***) |
| | | After PE | 0.76 (0.42) | | 0.77 (0.42) | | -3.81 (***) |
| | Passive off-task | Overall | 0.09 (0.29) | | 0.09 (0.28) | | 2.30 (*) |
| | | Before PE | 0.10 (0.30) | | 0.11 (0.31) | | -4.10 (***) |
| | | After PE | 0.08 (0.27) | | 0.07 (0.26) | | -3.03 (**) |
| | Active off-task | Overall | 0.13 (0.34) | | 0.14 (0.36) | | -4.06 (***) |
| | | Before PE | 0.14 (0.35) | | 0.14 (0.35) | | 6.88 (***) |
| | | After PE | 0.13 (0.34) | | 0.14 (0.08) | | -2.98 (**) |
| ID aggregated | On-task | Overall | 0.74 (0.10) | 0.46–0.97 | 0.74 (0.09) | 0.46–0.91 | -6.10 (***) |
| | Passive off-task | Overall | 0.09 (0.05) | 0.01-0.29 | 0.09 (0.06) | 0.01-0.29 | -8.04 (***) |
| | Active off-task | Overall | 0.13 (0.08) | 0.01–0.39 | 0.14 (0.08) | 0.02–0.39 | -17.39 (***) |
| | Other | Overall | 0.03 (0.03) | 0.00–0.17 | 0.03 (0.03) | 0.00–0.17 | 12.84 (***) |

Significance values *=p<0.05, **=p<0.01, ***=p<0.001

### Task descriptives

Task type was not evenly distributed across pre- and post-PE lessons in the analysed sub-sample $t_{(37,691.9)}$=-30.00, p<0.001. Pre-PE lessons mostly consisted of independent tasks, followed by teacher-led tasks. After PE this was the other way around, with teacher-led tasks more than twice as prevalent as before PE and independent task occurring less frequently than before PE (Table 12).

*Table 12. Distribution of task-type in pre- and post-PE lessons for the sub-sample with attainment scores.*

| Percentage occurrence of task type | Pre-PE | Post-PE |
|---|---|---|
| Independent task | 60.1% | 37.1% |
| Paired task | 1.1% | 3.5% |
| Group task | 2.1% | 2.2% |
| Teacher-led task | 18.1% | 41.6% |
| Other task | 18.6% | 15.5% |

### Control variables

Variables included as controls in the multilevel analyses were either at the inter-individual level (year group, BMI category, gender, teacher reports of attainment, and whether the participant participated in fewer or more minutes of organised sports per week than the sample mean), or at the within-lesson level (task type and self-report data). Descriptives of each of these control variables are presented, along with their correlation with on-task behaviour.

Within-lesson variables were derived from the self-reports completed during classroom and PE lessons. Participants (N=108) completed between 6 and 42 questionnaires during the intervention (M=37.4, SD=5.3, range 6-42), giving a total of 3,669 reports (3,093 classroom questionnaires, 576 PE questionnaires). 75% of the participants who filled out self-reports completed at least 24 of the 36 classroom questionnaires, and 94.4% completed at least 4 of the 6 PE questionnaires. The 76 participants who were observed and had their teacher report on their attainment levels completed

*Table 13. Self-report data descriptives for observed sample with attainment scores.*

| | Classroom (all lessons) | | | | | | | | PE | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | N | Min | Max | Mean | SD | Skewness | SE | Missing | N | Min | Max | Mean | SD | Skewness | SE | Missing |
| Task enjoyment | 2177 | 1 | 5 | 3.72 | 1.52 | -0.72 | 0.052 | 166 (7.1%) | | | | | | | | |
| Task difficulty | 2155 | 1 | 5 | 2.09 | 1.31 | 0.97 | 0.053 | 188 (8.0%) | | | | | | | | |
| Tiredness | 2171 | 1 | 5 | 2.98 | 1.49 | 0.05 | 0.053 | 172 (7.3%) | | | | | | | | |
| PE enjoyment | | | | | | | | | 568 | 1 | 5 | 4.00 | 1.31 | -1.14 | 0.052 | 8 (1.4%) |
| PE intensity | | | | | | | | | 569 | 1 | 5 | 3.24 | 1.50 | -0.22 | 0.052 | 7 (1.2%) |
| Excited | 2021 | 1 | 5 | 3.55 | 1.64 | -0.52 | 0.054 | 322 (13.7%) | 546 | 1 | 5 | 3.48 | 1.68 | -0.44 | 0.053 | 30 (5.2%) |
| Happy | 2018 | 1 | 5 | 3.92 | 1.43 | -0.97 | 0.054 | 325 (13.9%) | 545 | 1 | 5 | 4.01 | 1.38 | -1.10 | 0.054 | 31 (5.4%) |
| Calm | 2009 | 1 | 5 | 3.82 | 1.45 | -0.88 | 0.055 | 334 (14.3%) | 543 | 1 | 5 | 3.55 | 1.54 | -0.51 | 0.054 | 33 (5.7%) |
| Relaxed | 1998 | 1 | 5 | 3.84 | 1.46 | -0.89 | 0.055 | 345 (14.7%) | 537 | 1 | 5 | 3.59 | 1.55 | -0.59 | 0.054 | 39 (6.8%) |
| Bored | 1991 | 1 | 5 | 2.46 | 1.61 | 0.58 | 0.055 | 352 (15.0%) | 519 | 1 | 5 | 2.20 | 1.58 | 0.91 | 0.054 | 57 (9.9%) |
| Sad | 1889 | 1 | 5 | 1.53 | 1.14 | 2.22 | 0.056 | 454 (19.4%) | 502 | 1 | 5 | 1.652 | 1.13 | 2.27 | 0.055 | 74 (12.8%) |
| Angry | 1882 | 1 | 5 | 1.54 | 1.20 | 2.16 | 0.056 | 461 (19.7%) | 506 | 1 | 5 | 1.59 | 1.20 | 2.02 | 0.055 | 70 (12.2%) |
| Nervous | 1926 | 1 | 5 | 1.62 | 1.24 | 1.93 | 0.056 | 417 (17.8%) | 509 | 1 | 5 | 1.58 | 1.20 | 2.06 | 0.055 | 67 (11.6%) |

between twelve and 42 self-reports, and 80.3% completed at least 24 classroom reports. All but three of the participants in this subsample completed at least 4 PE reports. Descriptives of the self-report answers of the whole sample can be found in Table 13. The data was skewed for all self-reported items except tiredness, perceived PE intensity, and PE excitement. All positive emotions, task enjoyment, and PE enjoyment were negatively skewed, and negative emotion items and task difficulty were positively skewed. Skewness ranged between 0.58 and 2.27 for positively skewed items, and between -0.51 and -1.14 for negatively skewed items. Hereafter, data presented for the questionnaire answers will be exclusively for those items related to task/PE enjoyment, tiredness, and task difficulty. Emotion data was used in mediation analyses and is described in more detail in section 4.4.

Situation-specific task difficulty was rated 3 or below (easy to neutral) in 85.1% of reports, and was positively skewed. Reported task enjoyment was high in the classroom (M=3.72, SD=1.52), as well as in PE (M=4.00, SD=1.31): both PE and task enjoyment were negatively skewed. Classroom reports of tiredness were higher than expected; in 62.5% of the reports, participants reported feeling quite to very tired (3 or more stars out of 5), with 25.9% of reports at 5 out of 5 (Table 14). Perceived Rate of Exertion (PRE) after PE was rated 3 or more out of 5 in 70.8% of PE reports, and 5 out 5 in 34.8% of reports. This was not surprising, as 1/3 of the PE lessons were designed to be at high intensity.

*Table 14. Ratings of tiredness during classroom lessons.*

| Self-reported tiredness | | $N_{reports}$ |
|---|---|---|
| 1 star | 643 | (22.6%) |
| 2 stars | 422 | (14.9%) |
| 3 stars | 721 | (25.4%) |
| 4 stars | 319 | (11.2%) |
| 5 stars | 735 | (25.9%) |

Note: 1 = not at all tired, 5 = very tired.

Perceived task difficulty was positively related to on-task behaviour ($\tau_b$=0.016, p<0.01), and negatively to active off-task behaviour ($\tau_b$=-0.016, p<0.01), but not to passive off-task behaviour ($\tau_b$=-0.008, p=0.09). As it was related to on-task behaviour, perceived task difficulty was included as a control variable in the MLM.

Lesson content was not always consistent from week to week, and the proportion of on-task behaviour was examined in relation to participants' self-reported enjoyment. It was found to range between 0.69 and 0.79, and significantly relate to the proportion of time spent on-task ($\tau_b$=0.039, p<0.01) (Table 15). Therefore, task enjoyment was included as a control variable in the multilevel analyses.

*Table 15. Task behaviour by task enjoyment.*

| $N_{observations}$=36,342 | On-task | Passive off-task | Active off-task | Total observations | Tau b (p) |
|---|---|---|---|---|---|
| Task enjoyment | | | | | 0.039 (**) |
| 1 star | 0.70 (0.46) | 0.10 (0.30) | 0.17 (0.38) | 6,471 | |
| 2 stars | 0.69 (0.46) | 0.12 (0.32) | 0.16 (0.37) | 3,999 | |
| 3 stars | 0.74 0.44) | 0.09 (0.29) | 0.13 (0.34) | 8,614 | |
| 4 stars | 0.79 (0.41) | 0.07 (0.26) | 0.11 (0.31) | 4,556 | |
| 5 stars | 0.76 (0.43) | 0.09 (0.28) | 0.13 (0.33) | 23,982 | |

Note: Mean proportion of behaviour (SD). Significance values *=p<0.05, **=p<0.01, ***=p<0.001

*Table 16. Task behaviour by task type.*

| $N_{observations}$=38,933 | On-task | Passive off-task | Active off-task | Total observations | Tau b (p) |
|---|---|---|---|---|---|
| Task type | | | | | 0.058 (**) |
| Independent task | 0.72 (0.45) | 0.12 (0.32) | 0.15 (0.36) | 24,964 | |
| Paired task | 0.78 (0.42) | 0.08 (0.28) | 0.12 (0.33) | 1,215 | |
| Group task | 0.76 (0.43) | 0.08 (0.27) | 0.15 (0.36) | 1,115 | |
| Teacher-led task | 0.75 (0.43) | 0.09 (0.28) | 0.15 (0.36) | 15,551 | |
| Other task | 0.79 (0.40) | 0.03 (0.17) | 0.06 (0.24) | 8,795 | |

Note: Mean proportion of behaviour (SD). Significance values *=p<0.05, **=p<0.01, ***=p<0.001

As the occurrence of task type differed between the pre- and post-PE lesson, the proportion of on- and off-task behaviour was investigated for each task type (Table 16). Task type was significantly related to time on-task ($\tau_b$=0.058, p<0.01) and therefore deemed important to include as a control variable.

PE PRE was related to the intended lesson intensity ($\tau_b$=-0.254, p<0.01). Crosstabulation of PE PRE and intended lesson intensity showed that generally children perceived the high intensity lessons to be quite tiring, and low intensity lessons not to be very tiring, with medium intensity lessons in between (Table 17). Some ratings were quite high for the low intensity lessons, which may be a reflection of their fitness level. Self-reported tiredness in the classroom after PE was related to both PE PRE ($\tau_b$=-0.283, p<0.01) and the intended PE lesson intensity ($\tau_{b=}$-0.050, p<0.05). However, tiredness in the classroom was not related to on-task behaviour ($\tau_b$=-0.001, p=0.88) and thus not included in the multilevel analyses.

*Table 17. Crosstabulation of PE lesson design and PRE.*

| | Perceived Rate of Exertion | | | | |
|---|---|---|---|---|---|
| Lesson design | 1 | 2 | 3 | 4 | 5 |
| Low intensity | 252 | 155 | 204 | 123 | 207 |
| Medium intensity | 180 | 132 | 300 | 100 | 277 |
| High intensity | 66 | 84 | 201 | 153 | 489 |

Between-participant variables were investigated in relation to on-task behaviour. The proportions of observed task behaviours by the between-person variables year group, gender, and BMI is displayed in Table 18. Across the age groups there was significant difference in the proportion of on-task behaviour ($\tau_b$=-0.065, p<0.01), but the type of off-task behaviour differed. Whereas passive off-task behaviours decreased with age ($\tau_b$=-0.028, p<0.01), active off-task behaviours increased in prevalence ($\tau_b$=0.103, p<0.01). Girls exhibited more on-task behaviour than boys ($\tau_b$=0.039, p<0.01). Boys and girls

did not differ in their passive off-task behaviour ($\tau_b$=0.010, p=0.06), but girls displayed less active off-task behaviour during observed lessons ($\tau_b$=-0.053, p<0.01).

*Table 18. Task behaviour by year group, gender, and BMI.*

| $N_{observations}$=38,933 | On-task | Passive off-task | Active off-task | Total observations |
|---|---|---|---|---|
| Year 3 | 0.76 (0.43) | 0.12 (0.32) | 0.11 (0.31) | 6,252 |
| Year 4 | 0.77 (0.42) | 0.08 (0.28) | 0.11 (0.31) | 19,968 |
| Year 5 | 0.69 (0.46) | 0.08 (0.28) | 0.20 (0.40) | 12.713 |
| Male | 0.72 (0.45) | 0.09 (0.28) | 0.16 (0.37) | 15,792 |
| Female | 0.76 (0.43) | 0.09 (0.29) | 0.12 (0.33) | 23,141 |
| Low-to-healthy BMI | 0.76 (0.43) | 0.07 (0.26) | 0.14 (0.35) | 22,370 |
| Above-healthy BMI | 0.73 (0.45) | 0.12 (0.32) | 0.13 (0.34) | 16,563 |

Participants with an above-healthy BMI spent less time on-task ($\tau_b$=-0.034, p<0.01), more time passively off-task ($\tau_b$=0.078, p<0.01), and less time actively off-task ($\tau_b$=-0.019, p<0.01). All variables were significantly related to on-task behaviour and therefore included as control variables in the multilevel analyses.

On-task behaviour differed by attainment level ($\tau_b$=0.074, p<0.01), and attainment level was included as a control variable in the model. Those achieving above ARE displayed the greatest proportion of on-task behaviour, and those achieving below ARE the lowest (Table 19). In particular, passive off-task behaviour was lower in participants who achieved above ARE, with the proportion of passive off-task behaviour at half their proportion of active off-task behaviour.

*Table 19. Task behaviour by attainment.*

| $N_{observations}$=38,933 | On-task | Passive off-task | Active off-task | Total observations |
|---|---|---|---|---|
| Below ARE | 0.71 (0.45) | 0.12 (0.33) | 0.14 (0.35) | 5,557 |
| At ARE | 0.72 (0.45) | 0.10 (0.30) | 0.16 (0.36) | 18,360 |
| Above ARE | 0.79 (0.41) | 0.06 (0.24) | 0.12 (0.32) | 15,016 |

104 participants completed the 'trait PA questionnaire'. Fifteen participants (20.3%) did not take part in any sports except PE lessons, but the majority of the sample (69.2%) took part in 1-3 sports sessions a week. Mean participation in extra-curricular and out-of-school sport was 1.74 sessions per week (SD 1.45, range 0-6), and 108.18 minutes per week (SD 104.45, range 0-450). The dichotomous variable for participation in more sports sessions than the sample mean was allocated to 36 of the 76 participants in the sub-sample. Thirty participants equalled or exceeded the sample mean duration of sports participation.

The duration, but not the frequency, of sports participation was related to on-task behaviour ($\tau_{b(duration)}$=-0.023, p<0.01, $\tau_{b(frequency)}$=-0.002, p=0.67). Passive off-task behaviour was negatively related with sports participation at or above the sample mean for both duration ($\tau_b$=-0.020, p<0.01) and frequency ($\tau_b$=-0.032, p<0.01), whereas active off-task behaviour was positively related ($\tau_{b(duration)}$=0.032, p<0.01, $\tau_{b(frequency)}$=0.023, p<0.01). The dichotomous variable for duration of sports participation was included as a control variable in the MLM.

Objectively measured PA intensity was recorded over 24h on intervention days. Mean valid duration of the 24h recordings was 20.3h ($N_{recordings}$ 463, SD=2.4h, range 10.9h – 24.0h). The majority of missing data was during the night-time; daytime recordings (7am to 8pm) ranged between 8.83h and 13h of valid data (M=12.9h, SD=0.4h). The sample's mean MVPA duration over 24h, excluding PE lesson time, was 75.3 minutes (SD=31.0, range 21.0 – 226.0), and participants fell short of the government's recommendation of at least 60 minutes of MVPA per day on 35.6% of the recorded days. Participants who achieved government MVPA guidelines on at least 75% of the recorded days were called 'active kids'; this was 57.8% of the sample. Whether participants achieved the government's recommended amount of MVPA was cross-tabulated with their BMI category being 'healthy and below' or 'above healthy'. Participants with a healthy or low BMI failed to meet government guidelines on 32.5% of recorded days, compared to 40% for those with above-healthy BMI. Girls failed to meet the target on 46.7% of days, as opposed to boys on 18% of days. BMI status

did not significantly relate to acquiring the label 'active child', but gender did ($\tau_b$=-0.270, p<0.05). Being an 'active child' did not relate to on-task behaviour ($\tau_b$=0.007, p=0.23).

### 4.3.2 Multilevel Models of Physical Activity and Behavioural Engagement

Results for each step of the MLM building are provided in this section. First, the distribution of variance across the levels is presented. Then, the observation-level predictor PrePost (whether the observation takes place before or after PE) is added and results for the inclusion of random intercepts and slopes are presented. Next, research questions one to three are answered as the results of the predictor variables at the week-level (intervention effects) are presented. Finally, interactions of the intervention and between-participant variables are provided to answer research question four.

#### Null Model

The intercept-only model indicated variance at level 3 (participant) was 0.206 (p<0.001), level 2 variance (weeks) was 0.337 (p<0.001), and residual variance for logistic models is approximated to be $\pi^2/3$ = 3.29 (Goldstein *et al.*, 2002). The ICC for participants thus is:

$$ICC_{ID} = \frac{\rho_{ID}}{\rho_{residual} + \rho_{weeks} + \rho_{ID}} = \frac{0.206}{3.29 + 0.337 + 0.206} = 0.054$$

*Equation 25. Participant level ICC*

The ICC for weeks is:

$$ICC_{weeks} = \frac{\rho_{Weeks}}{\rho_{residual} + \rho_{weeks} + \rho_{ID}} = \frac{0.337}{3.29 + 0.337 + 0.206} = 0.088$$

*Equation 26. Week level ICC.*

This shows a greater amount of variation between weeks than between participants.

### Pre-/Post-PE On-Task Behaviour

The first model specified had PrePost added as the only predictor to investigate if on-task behaviour differed between lessons pre- and post-PE. The parameter for PrePost was positive and significant ($\beta_1$=0.213, SE=0.02, p<0.001) (Equation 27).

$$On-task_{ijk} = \beta_0 + \beta_1 PrePost_{ijk} + \nu_{0k} + \upsilon_{0jk}$$
$$= \mathbf{0.962} + \mathbf{0.213}\ PrePost_{ijk} + \nu_{0k} + \upsilon_{0jk}$$

*Equation 27. Model 1.1: Pre/Post PE behaviour – fixed effect.*

This model was first extended with a random intercept at the ID ($\nu_{0k}$) and week ($\upsilon_{0jk}$) level (Equation 28).

$$On-task_{ijk} = \beta_{0jk} + \beta_1 PrePost_{ijk}$$
$$\beta_{0jk} = \beta_0 + \nu_{0k} + \upsilon_{0jk}$$
$$On-task_{ijk} = \mathbf{0.951} + \mathbf{0.207}\ PrePost_{ijk} + \nu_{0k} + \upsilon_{0jk}$$
$$[\nu_{0k}] \sim N(0, \Omega): \Omega_\nu = [\sigma^2_{\nu 0}] = 0.204$$
$$[\upsilon_{0jk}] \sim N(0, \Omega_\upsilon): \Omega_\upsilon = [\sigma^2_{\upsilon 0}] = 0.335$$

*Equation 28. Model 1.2: Pre/Post PE behaviour – random intercept.*

Next, a random slope for PrePost was added to allow the strength of the relationship between PA and on-task behaviour to vary from week to week (Equation 29). Parameters for all three models are displayed in Table 20.

$$On-task_{ijk} = \beta_{0jk} + \beta_{1j} PrePost_{ijk}$$
$$\beta_{0jk} = \beta_0 + \nu_{0k} + \upsilon_{0jk}$$
$$\beta_{1j} = \beta_1 + \upsilon_{1jk}$$
$$On-task_{ijk} = \mathbf{0.959} + \mathbf{0.212}\ PrePost_{ijk} + \upsilon_{1jk} + \nu_{0k} + \upsilon_{0jk}$$
$$[\nu_{0k}] \sim N(0, \Omega): \Omega_\nu = [\sigma^2_{\nu 0}] = 0.170$$
$$\begin{bmatrix} \upsilon_{0jk} \\ \upsilon_{1jk} \end{bmatrix} \sim N(0, \Omega_\upsilon): \Omega_\upsilon = \begin{bmatrix} \sigma^2_{\upsilon 0} & \\ \sigma_{\upsilon 01} & \sigma^2_{\upsilon 1} \end{bmatrix} = \begin{bmatrix} 0.769 & \\ \mathbf{-0.731} & 1.137 \end{bmatrix}$$

*Equation 29. Model 1.3: Pre/Post PE behaviour – random intercept and random slope.*

*Table 20. Model 1.1-1.3: fixed and random parameters.*

| | 1.1 Fixed model | | | | | 1.2 Random intercept model | | | | | 1.3 Random intercept & slope model | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | β | S.E. | 95% CI | p | d | β | S.E. | 95% CI | p | d | β | S.E. | 95% CI | p | d |
| **Fixed Part** | | | | | | | | | | | | | | | |
| Intercept | 0.962 | 0.014 | 0.94-0.99 | *** | | 0.951 | 0.054 | 0.85-1.06 | *** | | 0.959 | 0.059 | 0.84-1.07 | *** | |
| PrePost | 0.213 | 0.020 | 0.17-0.25 | *** | 0.12 | 0.207 | 0.021 | 0.17-0.25 | *** | 0.11 | 0.212 | 0.051 | 0.11-0.31 | *** | 0.12 |
| **Random Part** | | | | | | | | | | | | | | | |
| Level: Week | | | | | | | | | | | | | | | |
| Intercept variance | | | | | | 0.335 | 0.026 | | *** | | 0.769 | 0.058 | | *** | |
| intercept/slope covariance | | | | | | | | | | | -0.731 | 0.062 | | *** | |
| Slope variance | | | | | | | | | | | 1.137 | 0.086 | | *** | |
| Level: ID | | | | | | | | | | | | | | | |
| Intercept variance | | | | | | 0.204 | 0.039 | | *** | | 0.170 | 0.034 | | *** | |
| Units: Week | 552 | | | | | 553 | | | | | 553 | | | | |
| Units: Obs_in_day | 51,640 | | | | | 51,640 | | | | | 51,640 | | | | |
| Units: ID | 101 | | | | | 101 | | | | | 101 | | | | |

Significance values *=p<0.05, **=p<0.01, ***=p<0.001

The eight control variables which were significantly related to on-task behaviour were added stepwise (starting with the most strongly correlated variable) and the effect size of the PrePost parameter was calculated at each step (Table 21). Full parameter estimates for each model with control variables are displayed in Table 22. All further models with predictors added were controlled for attainment, year group, task type, gender, task enjoyment, BMI category, duration of sports participation, and task difficulty.

*Table 21. Effect size of PrePost with stepwise addition of control variables.*

| Control variables included in the model | d |
|---|---|
| None | 0.12 |
| Attainment | 0.15 |
| Attainment & year group | 0.14 |
| Attainment, year group & task type | 0.14 |
| Attainment, year group, task type & gender | 0.15 |
| Attainment, year group, task type, gender & task enjoyment | 0.16 |
| Attainment, year group, task type, gender, task enjoyment & BMI category | 0.16 |
| Attainment, year group, task type, gender, task enjoyment, BMI category & duration of sports participation | 0.14 |
| Attainment, year group, task type, gender, task enjoyment, BMI category, duration of sports participation & task difficulty | 0.13 |

The fully controlled version of model 1 (model 1.4) found that on-task behaviour differed significantly between lesson pre- and post-PE, and that the slope for PrePost was random. The parameter for PrePost was significant ($\beta_{1j}$=0.243, SE=0.069, CI=0.11-0.38, p<0.001), though a Wald test for intercept + PrePost versus intercept only indicated that the difference in on-task behaviour between pre- and post-PE lessons was not significant ($\chi^2$=2.863, p=0.091). The effect size did not meet the cut-off for a small effect (d=0.13, OR=1.28). Table 22Table 20 and Equation 30 show the parameter estimates for model 1.4.

*Table 22. Stepwise addition of control variables in model 1.4.*

| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | β | S.E. | CI lower | CI upper | p | β | S.E. | CI lower | CI upper | p | β | S.E. | CI lower | CI upper | p |
| **Fixed Part** | | | | | | | | | | | | | | | |
| Intercept | 0.959 | 0.059 | 0.844 | 1.074 | *** | 0.810 | 0.081 | 0.650 | 0.970 | *** | 0.907 | 0.092 | 0.727 | 1.086 | *** |
| PrePost | 0.212 | 0.051 | 0.111 | 0.313 | *** | 0.269 | 0.061 | 0.149 | 0.388 | *** | 0.263 | 0.061 | 0.144 | 0.382 | *** |
| Well below ARE | | | | | | -0.036 | 0.221 | -0.469 | 0.398 | | -0.050 | 0.213 | -0.468 | 0.369 | |
| Below ARE | | | | | | -0.003 | 0.176 | -0.348 | 0.342 | | 0.016 | 0.174 | -0.325 | 0.358 | |
| Above ARE | | | | | | 0.189 | 0.122 | -0.050 | 0.427 | | 0.193 | 0.117 | -0.036 | 0.422 | |
| Well above ARE | | | | | | 0.644 | 0.152 | 0.345 | 0.943 | *** | 0.605 | 0.148 | 0.314 | 0.896 | *** |
| Year 3 | | | | | | | | | | | 0.027 | 0.144 | -0.254 | 0.309 | |
| Year 5 | | | | | | | | | | | -0.276 | 0.108 | -0.487 | -0.065 | * |
| **Random Part** | | | | | | | | | | | | | | | |
| Level: ID | | | | | | | | | | | | | | | |
| Intercept variance | 0.170 | 0.034 | | | *** | 0.115 | 0.030 | | | *** | 0.099 | 0.027 | | | *** |
| Level: Week | | | | | | | | | | | | | | | |
| Intercept variance | 0.769 | 0.058 | | | *** | 0.784 | 0.068 | | | *** | 0.763 | 0.066 | | | *** |

| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Intercept/slope covariance | -0.731 | 0.062 | | | *** | -0.785 | 0.075 | | | *** | -0.762 | 0.074 | | | *** |
| Slope variance | 1.137 | 0.086 | | | *** | 1.232 | 0.105 | | | *** | 1.213 | 0.104 | | | *** |
| Units: ID | 101 | | | | | 76 | | | | | 76 | | | | |
| Units: Week | 553 | | | | | 420 | | | | | 420 | | | | |
| Units: Observations | 51,640 | | | | | 38,933 | | | | | 38,933 | | | | |

ARE = Age related expectation. Ref. category: At ARE Ref. category: year 4

Significance values *=p<0.05, **=p<0.01, ***=p<0.001

| | β | S.E. | CI lower | CI upper | p | β | S.E. | CI lower | CI upper | p | β | S.E. | CI lower | CI upper | p |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Fixed Part | | | | | | | | | | | | | | | |
| Intercept | 0.777 | 0.094 | 0.592 | 0.962 | *** | 0.579 | 0.103 | 0.377 | 0.781 | *** | 0.582 | 0.116 | 0.355 | 0.809 | *** |
| PrePost | 0.258 | 0.063 | 0.135 | 0.382 | *** | 0.269 | 0.063 | 0.145 | 0.392 | *** | 0.286 | 0.066 | 0.157 | 0.415 | *** |
| Well below ARE | -0.056 | 0.218 | -0.483 | 0.371 | | -0.183 | 0.204 | -0.582 | 0.216 | | -0.106 | 0.211 | -0.519 | 0.307 | |
| Below ARE | 0.025 | 0.178 | -0.324 | 0.373 | | 0.046 | 0.163 | -0.272 | 0.365 | | 0.074 | 0.172 | -0.262 | 0.410 | |
| Above ARE | 0.188 | 0.119 | -0.046 | 0.422 | | 0.223 | 0.109 | 0.009 | 0.437 | * | 0.247 | 0.114 | 0.024 | 0.470 | * |
| Well above ARE | 0.596 | 0.151 | 0.300 | 0.893 | *** | 0.636 | 0.138 | 0.365 | 0.907 | *** | 0.656 | 0.146 | 0.369 | 0.942 | *** |
| Year 3 | 0.003 | 0.147 | -0.285 | 0.290 | | -0.017 | 0.134 | -0.280 | 0.246 | | -0.027 | 0.140 | -0.301 | 0.247 | |
| Year 5 | -0.300 | 0.110 | -0.516 | -0.084 | ** | -0.331 | 0.101 | -0.529 | -0.133 | ** | -0.307 | 0.106 | -0.515 | -0.100 | ** |

| | β | S.E. | CI lower | CI upper | p | β | S.E. | CI lower | CI upper | p | β | S.E. | CI lower | CI upper | p |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Paired task | 0.440 | 0.094 | 0.256 | 0.623 | *** | 0.447 | 0.094 | 0.263 | 0.630 | *** | 0.524 | 0.099 | 0.329 | 0.718 | *** |
| Group task | 0.185 | 0.102 | -0.015 | 0.384 | | 0.194 | 0.102 | -0.006 | 0.394 | | 0.234 | 0.106 | 0.027 | 0.442 | * |
| Teacher-led task | 0.085 | 0.041 | 0.004 | 0.165 | * | 0.087 | 0.041 | 0.007 | 0.168 | * | 0.121 | 0.043 | 0.037 | 0.206 | ** |
| Other | 0.590 | 0.039 | 0.514 | 0.666 | *** | 0.594 | 0.039 | 0.517 | 0.670 | *** | 0.681 | 0.041 | 0.601 | 0.761 | *** |
| Female | | | | | | 0.337 | 0.092 | 0.158 | 0.517 | *** | 0.284 | 0.096 | 0.096 | 0.473 | ** |
| I do not enjoy this task at all | | | | | | | | | | | -0.082 | 0.070 | -0.220 | 0.055 | |
| I do not really enjoy this task | | | | | | | | | | | -0.370 | 0.074 | -0.514 | -0.226 | *** |
| I like this task | | | | | | | | | | | 0.067 | 0.073 | -0.077 | 0.210 | |
| I like this task a lot | | | | | | | | | | | 0.020 | 0.059 | -0.097 | 0.136 | |
| Random Part | | | | | | | | | | | | | | | |
| Level: ID | | | | | | | | | | | | | | | |
| Intercept variance | 0.102 | 0.028 | | | *** | 0.075 | 0.024 | | | ** | 0.083 | 0.026 | | | ** |
| Level: Week | | | | | | | | | | | | | | | |
| Intercept variance | 0.817 | 0.070 | | | *** | 0.819 | 0.070 | | | *** | 0.797 | 0.070 | | | *** |
| Intercept/slope covariance | -0.818 | 0.079 | | | *** | -0.819 | 0.079 | | | *** | -0.816 | 0.081 | | | *** |
| Slope variance | 1.301 | 0.111 | | | *** | 1.294 | 0.110 | | | *** | 1.351 | 0.117 | | | *** |
| Units: ID | 76 | | | | | 76 | | | | | 76 | | | | |
| Units: Week | 420 | | | | | 420 | | | | | 411 | | | | |
| Units: Observations | 38,933 | | | | | 38,933 | | | | | 36,342 | | | | |

| | β | S.E. | CI lower | CI upper | p | β | S.E. | CI lower | CI upper | p | β | S.E. | CI lower | CI upper | p |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Ref. category: Independent task | | | | | Ref. category: Male | | | | | Ref. category: I do not like or dislike this task | | | | |

ARE = Age related expectation Significance values *=p<0.05, **=p<0.01, ***=p<0.001

| | β | S.E. | CI lower | CI upper | p | β | S.E. | CI lower | CI upper | p | β | S.E. | CI lower | CI upper | p |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Fixed Part** | | | | | | | | | | | | | | | |
| Intercept | 0.699 | 0.142 | 0.420 | 0.977 | *** | 0.747 | 0.141 | 0.471 | 1.023 | *** | 0.803 | 0.149 | 0.512 | 1.095 | *** |
| PrePost | 0.282 | 0.067 | 0.152 | 0.413 | *** | 0.256 | 0.067 | 0.124 | 0.387 | *** | 0.243 | 0.069 | 0.108 | 0.378 | *** |
| Well below ARE | -0.066 | 0.208 | -0.473 | 0.341 | | -0.073 | 0.202 | -0.469 | 0.324 | | -0.088 | 0.201 | -0.483 | 0.306 | |
| Below ARE | 0.111 | 0.170 | -0.223 | 0.444 | | 0.139 | 0.166 | -0.186 | 0.465 | | 0.114 | 0.180 | -0.239 | 0.468 | |
| Above ARE | 0.258 | 0.111 | 0.041 | 0.475 | * | 0.312 | 0.112 | 0.093 | 0.532 | ** | 0.308 | 0.112 | 0.089 | 0.526 | ** |
| Well above ARE | 0.657 | 0.142 | 0.378 | 0.936 | *** | 0.709 | 0.140 | 0.434 | 0.983 | *** | 0.719 | 0.139 | 0.445 | 0.992 | *** |
| Year 3 | -0.127 | 0.149 | -0.419 | 0.164 | | -0.105 | 0.147 | -0.393 | 0.183 | | -0.096 | 0.149 | -0.388 | 0.196 | |
| Year 5 | -0.360 | 0.108 | -0.573 | -0.148 | ** | -0.348 | 0.110 | -0.564 | -0.132 | ** | -0.337 | 0.110 | -0.552 | -0.122 | ** |
| Paired task | 0.525 | 0.099 | 0.331 | 0.720 | *** | 0.518 | 0.099 | 0.324 | 0.712 | *** | 0.490 | 0.100 | 0.293 | 0.686 | *** |
| Group task | 0.231 | 0.106 | 0.023 | 0.440 | * | 0.349 | 0.111 | 0.132 | 0.566 | ** | 0.354 | 0.113 | 0.132 | 0.576 | ** |

| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | β | S.E. | CI lower | CI upper | p | β | S.E. | CI lower | CI upper | p | β | S.E. | CI lower | CI upper | p |
| Teacher-led task | 0.121 | 0.043 | 0.037 | 0.206 | ** | 0.116 | 0.044 | 0.031 | 0.202 | ** | 0.113 | 0.045 | 0.026 | 0.200 | * |
| Other | 0.683 | 0.041 | 0.603 | 0.763 | *** | 0.665 | 0.041 | 0.584 | 0.746 | *** | 0.667 | 0.042 | 0.585 | 0.750 | *** |
| Female | 0.315 | 0.096 | 0.128 | 0.503 | ** | 0.313 | 0.096 | 0.125 | 0.502 | ** | 0.300 | 0.097 | 0.111 | 0.489 | ** |
| I do not enjoy this task at all | -0.077 | 0.070 | -0.215 | 0.060 | | -0.076 | 0.071 | -0.216 | 0.064 | | -0.087 | 0.075 | -0.233 | 0.059 | |
| I do not really enjoy this task | -0.368 | 0.074 | -0.513 | -0.224 | *** | -0.376 | 0.075 | -0.523 | -0.228 | *** | -0.399 | 0.076 | -0.547 | -0.250 | *** |
| I like this task | 0.069 | 0.074 | -0.076 | 0.213 | | 0.068 | 0.074 | -0.077 | 0.212 | | 0.081 | 0.075 | -0.066 | 0.227 | |
| I like this task a lot | 0.019 | 0.060 | -0.098 | 0.136 | | 0.024 | 0.060 | -0.094 | 0.141 | | 0.071 | 0.062 | -0.050 | 0.193 | |
| Below healthy BMI | -0.086 | 0.123 | -0.327 | 0.156 | | -0.039 | 0.126 | -0.285 | 0.207 | | -0.018 | 0.126 | -0.265 | 0.229 | |
| Above healthy BMI | -0.281 | 0.147 | -0.569 | 0.006 | | -0.264 | 0.143 | -0.544 | 0.016 | | -0.236 | 0.145 | -0.520 | 0.049 | |
| Overweight | -0.237 | 0.159 | -0.549 | 0.074 | | -0.200 | 0.155 | -0.504 | 0.104 | | -0.184 | 0.155 | -0.487 | 0.119 | |
| Obese | -0.079 | 0.149 | -0.371 | 0.213 | | -0.045 | 0.145 | -0.329 | 0.239 | | -0.039 | 0.147 | -0.327 | 0.249 | |
| Sports participation duration (min) ≥ sample mean | | | | | | -0.198 | 0.093 | -0.382 | -0.015 | * | -0.222 | 0.094 | -0.407 | -0.037 | * |
| This task is very easy | | | | | | | | | | | -0.120 | 0.061 | -0.240 | 0.000 | |
| This task is quite easy | | | | | | | | | | | -0.032 | 0.059 | -0.149 | 0.084 | |

| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | β | S.E. | CI lower | CI upper | p | β | S.E. | CI lower | CI upper | p | β | S.E. | CI lower | CI upper | p |
| This task is quite difficult | | | | | | | | | | | -0.013 | 0.092 | -0.193 | 0.168 | |
| This task is very difficult | | | | | | | | | | | -0.028 | 0.086 | -0.197 | 0.141 | |
| **Random Part** | | | | | | | | | | | | | | | |
| Level: ID | | | | | | | | | | | | | | | |
| Intercept variance | 0.071 | 0.024 | | | ** | 0.061 | 0.023 | | | ** | 0.059 | 0.023 | | | * |
| Level: Week | | | | | | | | | | | | | | | |
| Intercept variance | 0.798 | 0.070 | | | *** | 0.828 | 0.073 | | | *** | 0.834 | 0.075 | | | *** |
| Intercept/slope covariance | -0.821 | 0.081 | | | *** | -0.833 | 0.083 | | | *** | -0.851 | 0.086 | | | *** |
| Slope variance | 1.389 | 0.120 | | | *** | 1.368 | 0.121 | | | *** | 1.399 | 0.124 | | | *** |
| Units: ID | 76 | | | | | 74 | | | | | 73 | | | | |
| Units: Week | 411 | | | | | 400 | | | | | 394 | | | | |
| Units: Observations | 36,342 | | | | | 35,333 | | | | | 34,569 | | | | |
| | Ref. category: Healthy BMI | | | | | Ref. category: Participation duration < sample mean | | | | | Ref. category: This task is not easy or difficult | | | | |

ARE = Age related expectation Significance values *=p<0.05, **=p<0.01, ***=p<0.001

$$On-task_{ijk} = \beta_{0jk} + \beta_{1j} PrePost_{ijk} + \beta_{2-28} control\ variables$$

$$\beta_{0jk} = \beta_0 + \nu_{0k} + \upsilon_{0jk}$$

$$\beta_{1j} = \beta_1 + \upsilon_{1PrePost_{jk}}$$

$$On-task_{ijk} = \mathbf{0.765} + \mathbf{0.243}\ PrePost_{ijk} + \upsilon_{1PrePost_{jk}} + \nu_{0k} + \upsilon_{0jk}$$

$$[\nu_{0k}] \sim N(0, \Omega): \Omega_\nu = [\sigma^2_{\nu 0}] = 0.059$$

$$\begin{bmatrix} \upsilon_{0jk} \\ \upsilon_{1jk} \end{bmatrix} \sim N(0, \Omega_\upsilon): \Omega_\upsilon = \begin{bmatrix} \sigma^2_{\upsilon 0} & \\ \sigma_{\upsilon 01} & \sigma^2_{\upsilon 1} \end{bmatrix} = \begin{bmatrix} 0.834 & \\ \mathbf{-0.851} & 1.399 \end{bmatrix}$$

*Equation 30. Model 1.4: Pre/Post PE behaviour – random intercept and random slope with all control variables.*

Figure 24 shows the mean predicted probability of on-task behaviour before and after PE lessons, with 95% confidence intervals. It is clear that the mean for pre-PE on-task behaviour lies within the confidence interval of the post-PE lesson and vice versa, indicating a non-significant difference between the pre- and post-PE lessons.

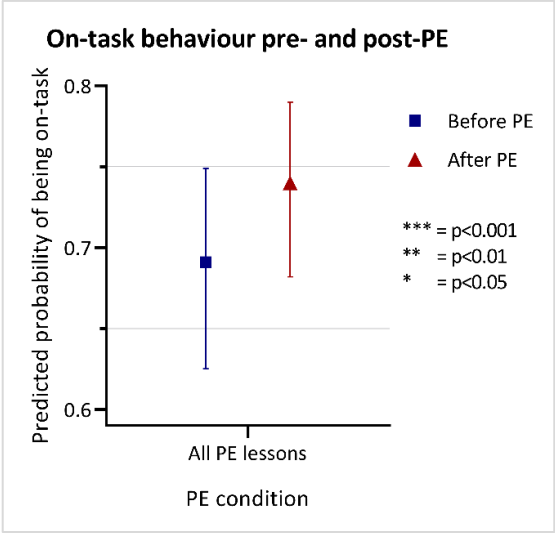



*Figure 24. Model 1.4: Mean and 95% confidence interval of on-task behaviour pre- and post-PE.*

The random slope-intercept covariance for PrePost ($\sigma_{u01}$= -0.851, SE=0.086, p<0.001) was negative and significant, indicating that for participants with a greater probability of on-task behaviour before PE, the improvement in on-task behaviour was lesser than in those with low probability of on-task behaviour before PE. This is illustrated in Figure 25: it is clearly visible here that those with lower on-task probability before PE have generally steeper slopes to their regression line than those with higher on-task probability before PE.

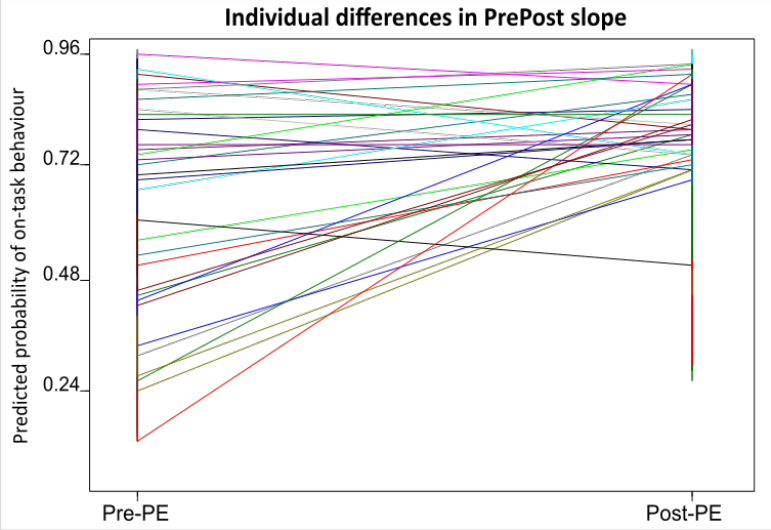



*Figure 25. Random intercept-slope covariance for a random sub-sample of participants.*

### RQ1. Main Effect of Complexity

Once PE lesson complexity was entered into the model as a predictor, estimates for the intercept and PrePost changed slightly. Complexity of the PE lesson did not significantly predict the proportion of on-task behaviour post-PE, and both the parameters for Complexity and for PrePost*Complexity were non-significant. The Wald test indicated no significant difference in pre- and post-PE on-task behaviour for low complexity or high complexity PE lessons ($\chi^2_{low\ complexity}$=1.835, p=0.176, $\chi^2_{high\ complexity}$=3.749, p=0.053). The random slope-intercept covariance for PrePost was again significant after the addition of Complexity to the model ($\sigma_{u01}$=-0.854, SE=0.086, p<0.001), meaning that those who were the least on-task before PE improved their behaviour the most after PE. Table 23 shows the parameter estimates for model 2, and these are used in Equation 31. Full parameter estimates including control variables for model 2 can be found in appendix J.

$$On-task_{ijk} = \beta_{0jk} + \beta_{1j}PrePost_{ijk} + \beta_2 Complexity_{jk} + \beta_3 PrePost * Complexity_{ijk} + \beta_{4-30} control\ variables$$

$$\beta_{0jk} = \beta_0 + \nu_{0k} + \upsilon_{0jk}$$

$$\beta_{1j} = \beta_1 + \upsilon_{1PrePost_{jk}}$$

$$On-task_{ijk} = \mathbf{0.786} + \mathbf{0.201}PrePost_{ijk} + \mathbf{0.038}Complexity_{jk} + \mathbf{0.089}PrePost * Complexity_{ijk} + \beta_{4-30}control\ variables + \upsilon_{1PrePost_{jk}} + \nu_{0k} + \upsilon_{0jk}$$

$$[\nu_{0k}] \sim N(0, \Omega): \Omega_\nu = [\sigma^2_{\nu 0}] = 0.059$$

$$\begin{bmatrix}\upsilon_{0jk}\\ \upsilon_{1jk}\end{bmatrix} \sim N(0, \Omega_\upsilon): \Omega_\upsilon = \begin{bmatrix}\sigma^2_{\upsilon 0} & \\ \sigma_{\upsilon 01} & \sigma^2_{\upsilon 1}\end{bmatrix} = \begin{bmatrix}0.835 & \\ \mathbf{-0.854} & 1.398\end{bmatrix}$$

*Equation 31. Model 2: PrePost-by-Complexity.*

*Table 23. Model 2: PrePost-by-Complexity: parameter estimates*

| | β | SE | CI lower | CI upper | p |
|---|---|---|---|---|---|
| **Fixed part** | | | | | |
| Intercept | 0.786 | 0.159 | 0.475 | 1.098 | *** |
| PrePost | 0.201 | 0.098 | 0.009 | 0.393 | * |
| Complexity | 0.038 | 0.105 | -0.168 | 0.243 | |
| PrePost * Complexity | 0.089 | 0.135 | -0.176 | 0.354 | |
| **Random part** | | | | | |
| Level: ID | | | | | |
| Intercept variance | 0.059 | 0.023 | | | * |
| Level: Week | | | | | |
| Intercept variance | 0.835 | 0.075 | | | *** |
| Intercept/slope covariance | -0.854 | 0.086 | | | *** |
| Slope variance | 1.398 | 0.124 | | | *** |
| Units: ID | 73 | | | | |
| Units: Week | 394 | | | | |
| Units: Observations | 34,569 | | | | |

Significance values *=p<0.05, **=p<0.01, ***=p<0.001

As can be seen in Figure 26, the increase in probability of on-task behaviour after PE was greater after high complexity PE than after low complexity PE, but this difference still was not significant. Effect sizes for both low and high complexity lessons were similar (Figure 27), and neither low- nor high-complexity lessons affected subsequent on-task behaviour with an effect size of greater than d=0.2 (low complexity: d=0.11, OR=1.22; high complexity: d=0.16, OR=1.34). Thus, no main effect of complexity was found, and the hypothesis that exercise complexity has a main effect on on-task behaviour was rejected.

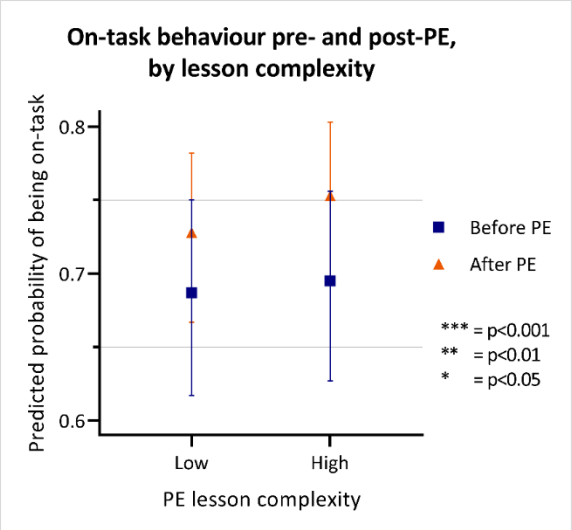



*Figure 26. Model 2: Mean and 95% confidence interval of on-task behaviour pre- and post-PE by complexity.*

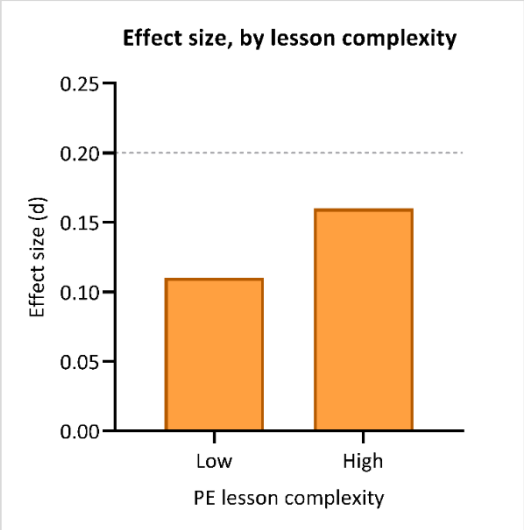



*Figure 27. Model 2: Effect size of PE lesson complexity.*

### RQ2. Main Effect of Intensity

In model 3, PE lesson intensity was entered into the model instead of complexity. No parameters in this model were significant in isolation. However, Wald tests showed that when all parameters applicable to each pre- and post-PE lesson were analysed together at each intensity level, post-PE on-task behaviour was significantly higher after high-intensity PE lessons, but not after low- or medium-intensity PE lessons ($\chi^2_{low\ intensity}$=1.594, p=0.207; $\chi^2_{medium\ intensity}$=1.629, p=0.202; $\chi^2_{high\ intensity}$=4.834, p<0.05). Parameter estimates for model 3 can be found in Table 24 and Equation 32.

$$On-task_{ijk} = \beta_{0jk} + \beta_{1j}PrePost_{ijk} + \beta_2 Intensity_medium_{jk} + \beta_3 Intensity_high_{jk} + \beta_4 PrePost * Intensity_medium_{ijk} + \beta_5 PrePost * Intensity_medium_{ijk} + \beta_{6-32} control\ variables$$

$$\beta_{0jk} = \beta_0 + \nu_{0k} + \upsilon_{0jk}$$

$$\beta_{1j} = \beta_1 + \upsilon_{1PrePost_{jk}}$$

$$On-task_{ijk} = \mathbf{0.865} + \mathbf{0.196}PrePost_{ijk} + \mathbf{-0.152}Intensity_{medium\,jk} + \mathbf{-0.027}Intensity_{high\,jk} + \mathbf{0.001}PrePost * Intensity_{medium\,ijk} + \mathbf{0.145}PrePost * Intensity_medium_{ijk} + \beta_{6-32}control\ variables + \upsilon_{1PrePost_{jk}} + \nu_{0k} + \upsilon_{0jk}$$

$$[\nu_{0k}] \sim N(0, \Omega): \Omega_\nu = [\sigma^2_{\nu 0}] = 0.062$$

$$\begin{bmatrix}\upsilon_{0jk}\\ \upsilon_{1jk}\end{bmatrix} \sim N(0, \Omega_\upsilon): \Omega_\upsilon = \begin{bmatrix}\sigma^2_{\upsilon 0} & \\ \sigma_{\upsilon 01} & \sigma^2_{\upsilon 1}\end{bmatrix} = \begin{bmatrix}0.835 & \\ \mathbf{-0.862} & 1.397\end{bmatrix}$$

*Equation 32. Model 3: PrePost-by-Intensity.*

*Table 24. Model 3: PrePost-by-Intensity: parameter estimates.*

| | β | SE | CI lower | CI upper | p |
|---|---|---|---|---|---|
| **Fixed part** | | | | | |
| Intercept | 0.865 | 0.166 | 0.540 | 1.190 | *** |
| PrePost | 0.196 | 0.119 | -0.037 | 0.430 | |
| Intensity_medium | -0.152 | 0.128 | -0.404 | 0.099 | |
| Intensity_high | -0.027 | 0.130 | -0.281 | 0.227 | |
| PrePost * Intensity_medium | 0.001 | 0.165 | -0.324 | 0.325 | |
| PrePost * Intensity_high | 0.145 | 0.166 | -0.181 | 0.471 | |
| **Random part** | | | | | |
| Level: ID | | | | | |
| Intercept variance | 0.062 | 0.023 | | | ** |
| Level: Week | | | | | |
| Intercept variance | 0.835 | 0.075 | | | *** |
| Intercept/slope covariance | -0.862 | 0.086 | | | *** |
| Slope variance | 1.397 | 0.124 | | | *** |
| Units: ID | 73 | | | | |
| Units: Week | 394 | | | | |
| Units: Observations | 34,569 | | | | |

Significance values *=p<0.05, **=p<0.01, ***=p<0.001

The slope-intercept covariance for PrePost again remained negative and significant ($\sigma_{u01}$=-0.862, SE=0.086, p<0.001), indicating that those who were less on-task before PE benefited more strongly

from PE lessons. Full parameter estimates including control variables for model 3 can be found in appendix K.

Figure 28 shows the predicted probability of on-task behaviour pre- and post-PE for all three intensity levels. The difference between the pre- and post-PE lesson was greatest for high intensity PE lessons, which is reflected in the effect sizes (low intensity: d=0.11, OR=1.22; medium intensity: d=0.11, OR=1.22; high intensity: d=0.19, OR=1.41, Figure 29). Although high intensity had a significant effect on post-PE on-task behaviour, the effect size still did not exceed d=0.2, and the hypothesis that there is a main effect of PA intensity on on-task behaviour was rejected for all levels of PE lesson intensity.

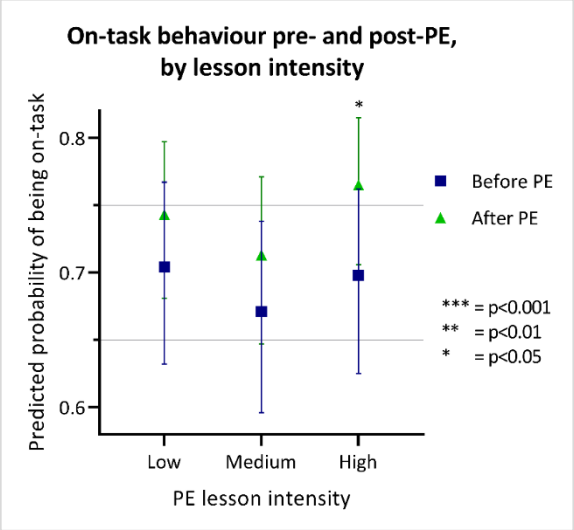



*Figure 28. Model 3: Mean and 95% confidence interval of on-task behaviour pre- and post-PE by intensity.*

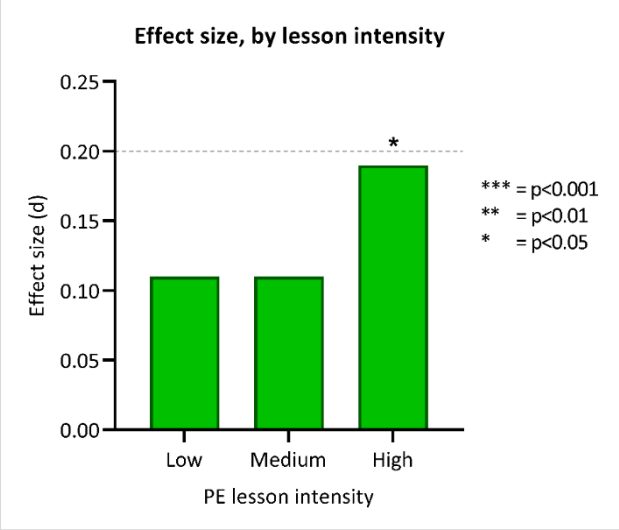



*Figure 29. Model 3: Effect size of PE lesson intensity.*

### RQ3. Interaction Effect of Complexity and Intensity

Finally, PE lesson complexity and intensity were both added to the model. In this model, the post-PE probability of on-task behaviour was not significantly predicted by any of the parameter estimates in isolation. Wald test results showed a significant difference in pre- and post-PE on-task behaviour for high intensity-high complexity PE lessons, but not other PE conditions ($\chi^2_{low-low}$=0.915, p=0.339; $\chi^2_{medium-low}$=3.850, p=0.050; $\chi^2_{high-low}$=0.283, p=0.594; $\chi^2_{low-high}$=1.668, p=0.197; $\chi^2_{medium-high}$=0.170, p=0.680; $\chi^2_{high-high}$=12.197, p<0.001). Table 25 and Equation 33 display the parameter estimates for the predictor variables, and the random terms of model 4.

$$
\begin{aligned}
On-task_{ijk} = {} & \beta_{0jk} + \beta_{1j}PrePost_{ijk} + \beta_2 Medium-Low_{jk} + \beta_3 High-Low_{jk} \\
& + \beta_4 Low-High_{jk} + \beta_5 Medium-High_{jk} + \beta_6 High-High_{jk} \\
& + \beta_7 PrePost * Medium-Low_{ijk} + \beta_8 PrePost * High-Low_{ijk} \\
& + \beta_9 PrePost * Low-High_{ijk} + \beta_{10} PrePost * Medium-High_{ijk} \\
& + \beta_{11} PrePost * High-High_{ijk} + \beta_{12-38} control\ variables
\end{aligned}
$$

$$\beta_{0jk} = \beta_0 + \nu_{0k} + \upsilon_{0jk}$$

$$\beta_{1j} = \beta_1 + \upsilon_{1PrePost_{jk}}$$

$$
\begin{aligned}
On-task_{ijk} = {} & \mathbf{0.759} + \mathbf{0.164} PrePost_{ijk} + \mathbf{-0.066} Medium-Low_{jk} + \mathbf{0.142} High-Low_{jk} \\
& + \mathbf{0.213} Low-High_{jk} + \mathbf{-0.038} Medium-High_{jk} + \mathbf{0.024} High-High_{jk} \\
& + \mathbf{0.169} PrePost * Medium-Low_{ijk} + \mathbf{-0.074} PrePost * High-Low_{ijk} \\
& + \mathbf{0.059} PrePost * Low-High_{ijk} + \mathbf{-0.094} PrePost * Medium-High_{ijk} \\
& + \mathbf{0.436} PrePost * High-High_{ijk} + \beta_{12-38} control\ variables + \upsilon_{1PrePost_{jk}} \\
& + \nu_{0k} + \upsilon_{0jk}
\end{aligned}
$$

$$[\nu_{0k}] \sim N(0, \Omega): \Omega_\nu = [\sigma^2_{\nu 0}] = 0.065$$

$$\begin{bmatrix} \upsilon_{0jk} \\ \upsilon_{1jk} \end{bmatrix} \sim N(0, \Omega_\upsilon): \Omega_\upsilon = \begin{bmatrix} \sigma^2_{\upsilon 0} & \\ \sigma_{\upsilon 01} & \sigma^2_{\upsilon 1} \end{bmatrix} = \begin{bmatrix} 0.832 & \\ \mathbf{-0.861} & 1.370 \end{bmatrix}$$

*Equation 33. Model 4: PrePost-by-Complexity-by-Intensity.*

*Table 25. Model 4: Interaction of intensity-by-complexity - parameter estimates.*

| | β | S.E. | CI lower | CI upper | p |
|---|---|---|---|---|---|
| **Fixed Part** | | | | | |
| Intercept | 0.759 | 0.19 | 0.386 | 1.132 | *** |
| PrePost | 0.164 | 0.166 | -0.161 | 0.488 | |
| Medium intensity, low complexity | -0.066 | 0.184 | -0.426 | 0.295 | |
| High intensity, low complexity | 0.142 | 0.187 | -0.224 | 0.508 | |
| Low intensity, high complexity | 0.213 | 0.183 | -0.146 | 0.572 | |
| Medium intensity, high complexity | -0.038 | 0.176 | -0.383 | 0.308 | |
| High intensity, high complexity | 0.024 | 0.176 | -0.321 | 0.369 | |
| PrePost*Medium intensity, low complexity | 0.169 | 0.235 | -0.291 | 0.63 | |
| PrePost*High intensity, low complexity | -0.074 | 0.237 | -0.538 | 0.39 | |
| PrePost*Low intensity, high complexity | 0.059 | 0.236 | -0.403 | 0.52 | |
| PrePost*Medium intensity, high complexity | -0.094 | 0.229 | -0.542 | 0.354 | |
| PrePost*High intensity, high complexity | 0.436 | 0.23 | -0.014 | 0.887 | |
| **Random Part** | | | | | |
| Level: ID | | | | | |
| Intercept variance | 0.065 | 0.023 | | | ** |
| Level: Week | | | | | |
| Intercept variance | 0.832 | 0.074 | | | *** |
| Intercept/slope covariance | -0.861 | 0.085 | | | *** |
| Slope variance | 1.370 | 0.122 | | | *** |
| Units: ID | 73 | | | | |
| Units: Week | 394 | | | | |
| Units: Observations | 34569 | | | | |

Significance values *=p<0.05, **=p<0.01, ***=p<0.001

The random slope-intercept covariance was once more negative and significant in this model ($\sigma_{v01}$=-0.861, SE=0.085, p<0.001), which means that those with the lowest on-task behaviour before PE gained the most from PE lessons. Full parameter estimates including control variables for model 4 can be found in appendix L.

Figure 30 shows the predicted probability of on-task behaviour pre- and post-PE for all intervention conditions. It shows that although post-PE the predicted probability of on-task behaviour is greater, this difference is not significant in five of the conditions as the confidence intervals have large

overlaps. There is however also a clear difference in the extent of overlap in confidence intervals between pre- and post-PE lessons.

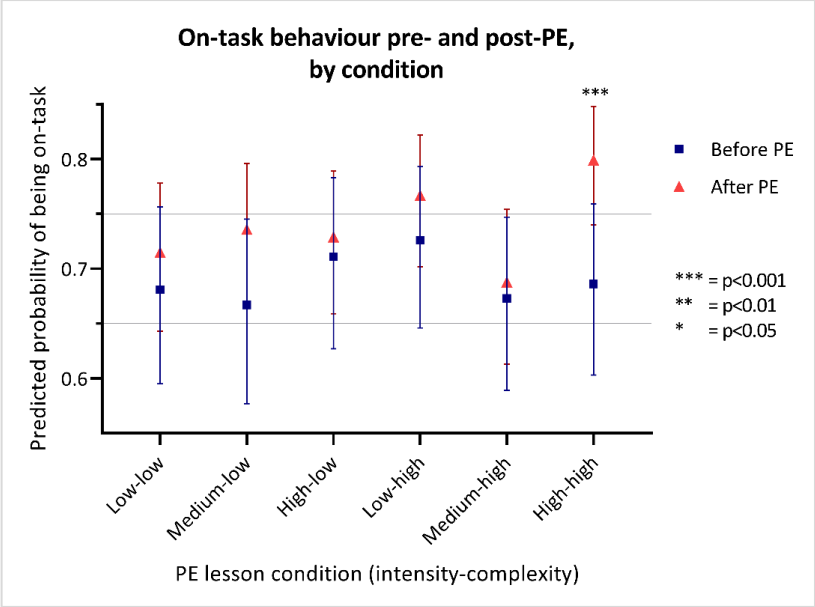



*Figure 30. Model 4: Mean and 95% confidence interval of on-task behaviour pre- and post-PE by complexity-by-intensity.*

When effect sizes were calculated it was found that complexity and intensity did interact, as the dose-response relationship for intensity and on-task behaviour is different at low complexity than at high complexity. A small positive effect size was found for high intensity with high complexity (d=0.33, OR=1.82). This was the only condition to exceed the d=0.2 cut-off. Next were medium intensity/low complexity (d=0.18, OR=1.40), and low intensity/high complexity lessons (d=0.12, OR=1.25). The other three conditions had effect sizes below d=0.1 (low intensity/low complexity: d=0.09, OR=1.18; high intensity/low complexity: d=0.05, OR=1.10; medium intensity/high complexity: d=0.04, OR=1.07). In terms of the difference in probability of on-task behaviour occurring between pre- and post-PE lessons, an effect size of d=0.33 reflects pupils being 16% more likely to be on-task after PE than before PE, and an effect size of d=0.18 corresponds to a 10% increase in probability.

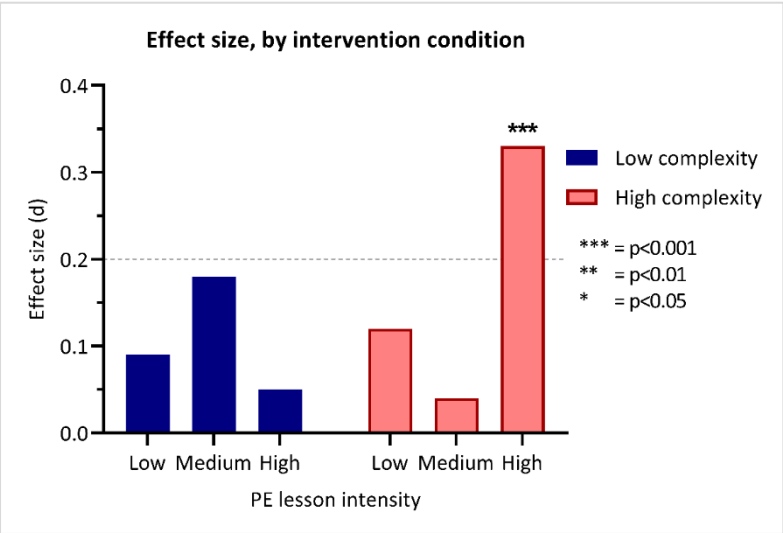



*Figure 31. Model 4: Effect size of PE lesson complexity-by-intensity.*

Figure 31 displays the effect size of each condition. Effect sizes differed between conditions; low complexity

lessons showed a slight inverted-U for intensity, but high complexity lessons did not. A U-relationship, with greater effects of low and high intensity exercise was observed in high complexity lessons. Therefore, the hypothesis that there is an interaction effect for PA complexity and intensity on on-task behaviour was accepted, although the hypothesised direction of the interaction effect was rejected for high complexity PA.

### RQ4. Individual Differences in the Response to Physical Activity

ID-level variables were added to the model specification used to answer RQ 4. These were entered as predictors, and as interactions with PrePost, Intensity, and Complexity. First, the effect of PA on children who were below the 33rd percentile for on-task behaviour before PE was examined (model 5.1). Then, teacher reports of attainment (trichotomized to below ARE, at ARE, and above ARE) were entered (model 5.2). In model 5.3, the effect of weekly duration of sports participation was inspected. Finally, the difference in PA effects by gender were considered (model 5.4). The four variables included in models 5.1-5.4 to investigate if the effect of PA differed for certain groups of participants did not correlate with each other.

*Pre-PE on-task behaviour.* Wald tests found that those who were in the bottom third of on-task behaviour before PE showed significant improvements in on-task behaviour post-PE for all PE lesson conditions ($\chi^2_{low-low}$=19.350, p<0.001; $\chi^2_{medium-low}$=18.727, p<0.001; $\chi^2_{high-low}$=6.588, p<0.05; $\chi^2_{low-high}$=24.290, p<0.001; $\chi^2_{medium-high}$=12.912, p<0.001; $\chi^2_{high-high}$=35.566, p<0.001). The sample as a whole, in contrast, only significantly increased their probability of on-task behaviour after high intensity, high complexity PE lessons. Parameter estimates for the predictor variables in model 5.1 can be found in Table 26 and Equation 34. Full parameter estimates are included in appendix M.

*Table 26. Model 5.1: Intensity-by-complexity-by-pre-PE on-task - parameter estimates.*

| | β | S.E. | CI lower | CI upper | p |
|---|---|---|---|---|---|
| **Fixed Part** | | | | | |
| Intercept | 1.290 | 0.192 | 0.914 | 1.666 | *** |
| PrePost | -0.325 | 0.191 | -0.699 | 0.050 | |
| Medium intensity, low complexity | -0.260 | 0.214 | -0.680 | 0.159 | |
| High intensity, low complexity | 0.095 | 0.221 | -0.338 | 0.529 | |
| Low intensity, high complexity | 0.195 | 0.211 | -0.218 | 0.609 | |
| Medium intensity, high complexity | -0.056 | 0.204 | -0.457 | 0.344 | |
| High intensity, high complexity | -0.197 | 0.201 | -0.591 | 0.198 | |
| PrePost*Medium intensity, low complexity | 0.333 | 0.273 | -0.202 | 0.869 | |
| PrePost*High intensity, low complexity | 0.085 | 0.277 | -0.459 | 0.628 | |
| PrePost*Low intensity, high complexity | 0.004 | 0.271 | -0.528 | 0.535 | |
| PrePost*Medium intensity, high complexity | -0.014 | 0.264 | -0.532 | 0.504 | |
| PrePost*High intensity, high complexity | 0.594 | 0.263 | 0.078 | 1.110 | * |
| Below 33rd percentile on-task | -1.332 | 0.259 | -1.840 | -0.823 | *** |
| PrePost*Below 33rd percentile on-task | 1.271 | 0.329 | 0.626 | 1.917 | *** |
| Medium intensity, low complexity* Below 33rd percentile on-task | 0.439 | 0.361 | -0.268 | 1.147 | |
| High intensity, low complexity* Below 33rd percentile on-task | 0.113 | 0.362 | -0.598 | 0.823 | |
| Low intensity, high complexity* Below 33rd percentile on-task | 0.028 | 0.366 | -0.689 | 0.745 | |
| Medium intensity, high complexity* Below 33rd percentile on-task | -0.063 | 0.349 | -0.746 | 0.621 | |
| High intensity, high complexity* Below 33rd percentile on-task | 0.391 | 0.353 | -0.300 | 1.082 | |
| PrePost*Medium intensity, low complexity* Below 33rd percentile on-task | -0.375 | 0.464 | -1.284 | 0.534 | |
| PrePost*High intensity, low complexity* Below 33rd percentile on-task | -0.492 | 0.464 | -1.402 | 0.417 | |
| PrePost*Low intensity, high complexity* Below 33rd percentile on-task | 0.124 | 0.471 | -0.799 | 1.047 | |
| PrePost*Medium intensity, high complexity* Below 33rd percentile on-task | -0.176 | 0.454 | -1.065 | 0.713 | |
| PrePost*High intensity, high complexity* Below 33rd percentile on-task | -0.233 | 0.461 | -1.137 | 0.670 | |

| | β | S.E. | CI lower | CI upper | p |
|---|---|---|---|---|---|
| **Random Part** | | | | | |
| Level: ID | | | | | |
| Intercept variance | 0.020 | 0.016 | | | |
| Level: Week | | | | | |
| Intercept variance | 0.707 | 0.065 | | | *** |
| Intercept/slope covariance | -0.692 | 0.074 | | | *** |
| Slope variance | 1.165 | 0.108 | | | *** |
| Units: ID | 73 | | | | |
| Units: Week | 394 | | | | |
| Units: Observations | 34,569 | | | | |

Significance values *=p<0.05, **=p<0.01, ***=p<0.001

$$
\begin{aligned}
On-task_{ijk} = {} & \beta_{0jk} + \beta_{1j}PrePost_{ijk} + \beta_2 Medium-Low_{jk} + \beta_3 High-Low_{jk} \\
& + \beta_4 Low-High_{jk} + \beta_5 Medium-High_{jk} + \beta_6 High-High_{jk} \\
& + \beta_7 PrePost * Medium-Low_{ijk} + \beta_8 PrePost * High-Low_{ijk} \\
& + \beta_9 PrePost * Low-High_{ijk} + \beta_{10} PrePost * Medium-High_{ijk} \\
& + \beta_{11} PrePost * High-High_{ijk} + \beta_{12} Below\ 33rd\ percentile_k \\
& + \beta_{13} PrePost * Below\ 33rd\ percentile_{ijk} \\
& + \beta_{14} Medium-Low * Below\ 33rd\ percentile_{jk} \\
& + \beta_{15} High-Low * Below\ 33rd\ percentile_{jk} \\
& + \beta_{16} Low-High * Below\ 33rd\ percentile_{jk} \\
& + \beta_{17} Medium-High * Below\ 33rd\ percentile_{jk} \\
& + \beta_{18} High-High * Below\ 33rd\ percentile_{jk} \\
& + \beta_{19} PrePost * Medium-Low * Below\ 33rd\ percentile_{ijk} \\
& + \beta_{20} PrePost * High-Low * Below\ 33rd\ percentile_{ijk} \\
& + \beta_{21} PrePost * Low-High * Below\ 33rd\ percentile_{ijk} \\
& + \beta_{22} PrePost * Medium-High * Below\ 33rd\ percentile_{ijk} \\
& + \beta_{23} PrePost * High-High * Below\ 33rd\ percentile_{ijk} \\
& + \beta_{24-47} control\ variables
\end{aligned}
$$

$$\beta_{0jk} = \beta_0 + \nu_{0k} + \upsilon_{1PrePost_{jk}}$$

$$\beta_{1j} = \beta_1 + \upsilon_{1jk}$$

$$
\begin{aligned}
On-task_{ijk} = {} & \mathbf{1.290} + \mathbf{-0.325}PrePost_{ijk} + \mathbf{-0.260}Medium-Low_{jk} \\
& + \mathbf{-0.095}High-Low_{jk} + \mathbf{-0.195}Low-High_{jk} + \mathbf{-0.056}Medium-High_{jk} \\
& + \mathbf{-0.197}High-High_{jk} + \mathbf{0.333}PrePost*Medium-Low_{ijk} \\
& + \mathbf{0.085}PrePost*High-Low_{ijk} + \mathbf{-0.004}PrePost*Low-High_{ijk} \\
& + \mathbf{-0.014}PrePost*Medium-High_{ijk} + \mathbf{0.594}PrePost*High-High_{ijk} \\
& + \mathbf{-1.332}Below\ 33rd\ percentile_{k} \\
& + \mathbf{1.271}PrePost*Below\ 33rd\ percentile_{ijk} \\
& + \mathbf{0.439}Medium-Low*Below\ 33rd\ percentile_{jk} \\
& + \mathbf{0.113}High-Low*Below\ 33rd\ percentile_{jk} \\
& + \mathbf{0.028}Low-High*Below\ 33rd\ percentile_{jk} \\
& + \mathbf{-0.063}Medium-High*Below\ 33rd\ percentile_{jk} \\
& + \mathbf{0.391}High-High*Below\ 33rd\ percentile_{jk} \\
& + \mathbf{-0.375}PrePost*Medium-Low*Below\ 33rd\ percentile_{ijk} \\
& + \mathbf{-0.492}PrePost*High-Low*Below\ 33rd\ percentile_{ijk} \\
& + \mathbf{0.124}PrePost*Low-High*Below\ 33rd\ percentile_{ijk} \\
& + \mathbf{-0.176}PrePost*Medium-High*Below\ 33rd\ percentile_{ijk} \\
& + \mathbf{-0.233}PrePost*High-High*Below\ 33rd\ percentile_{ijk} \\
& + \beta_{24-47}control\ variables + \upsilon_{1PrePost_{jk}} + \nu_{0k} + \upsilon_{0jk}
\end{aligned}
$$

$$[\nu_{0k}] \sim N(0,\Omega): \Omega_{\nu} = [\sigma^2_{\nu 0}] = 0.020$$

$$\begin{bmatrix}\upsilon_{0jk}\\ \upsilon_{1jk}\end{bmatrix} \sim N(0,\Omega_{\upsilon}): \Omega_{\upsilon} = \begin{bmatrix}\sigma^2_{\upsilon 0} & \\ \sigma_{\upsilon 01} & \sigma^2_{\upsilon 1}\end{bmatrix} = \begin{bmatrix}0.707 & \\ \mathbf{-0.692} & 1.165\end{bmatrix}$$

*Equation 34. Model 5.1. Intensity-by-complexity-by-pre-PE on-task behaviour.*

Effect sizes for all PE lesson conditions were greater for those participants who were the least on-task before PE (Figure 32). All conditions had effect sizes greater than 0.2, and four conditions had medium effect sizes (d≥0.5.). The greatest effect size was still for high intensity, high complexity (d=0.72, OR=3.70). The random intercept at the ID level was no longer significant in this model. The hypothesis that those with lowest levels of on-task behaviour would benefit the most from participation in PA was accepted for all intervention conditions.

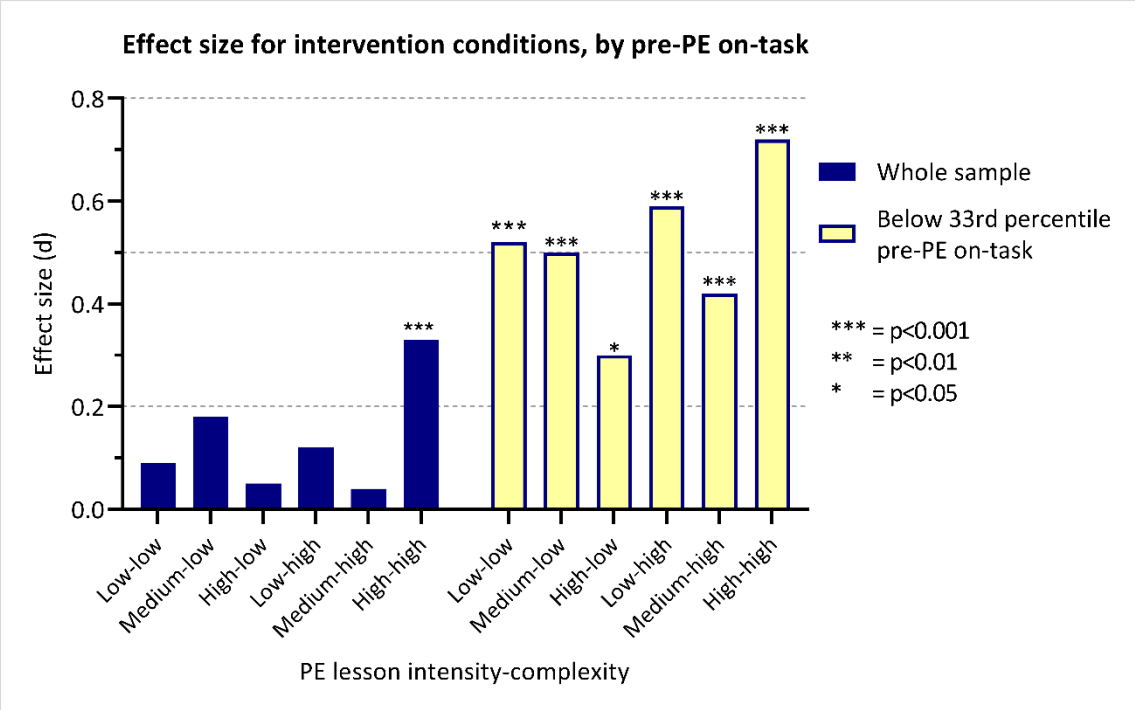



*Figure 32. Effect size of PE lesson intensity-by-complexity for the whole sample and those below the 33rd percentile on-task.*

*Attainment.* Participants were not equally distributed across the three categories of attainment; twelve participants were rated 'below ARE', 34 were rated 'at ARE', and 30 were rated 'above ARE'. Participants' response to PA differed by teacher-reported attainment level. Results of the Wald test showed that those in the 'below ARE' category were significantly more on-task post-PE for one of the PE lesson conditions (high-high), compared to two conditions for those 'at ARE' (medium-low and high-high), and one condition for those 'above ARE' (low-high) (Table 27). Parameter estimates for the predictor variables in model 5.2 can be found in Table 28 and Equation 35. Full parameter estimates are included in appendix N.

*Table 27. Chi^2 for each intervention condition, split by teacher-rated attainment.*

| | Low complexity | | | High complexity | | |
|---|---|---|---|---|---|---|
| | Low | Medium | High | Low | Medium | High |
| Below ARE | 1.793 | 0.943 | 2.117 | 1.673 | 0.844 | 9.379 (**) |
| At ARE | 0.170 | 6.064 (*) | 3.471 | 1.406 | 1.757 | 12.261 (***) |
| Above ARE | 0.591 | 0.855 | 4.596 (*) | 0.624 | 2.811 | 3.635 |

Significance values *=p<0.05, **=p<0.01, ***=p<0.001

*Table 28. Model 5.2: Intensity-by-complexity-by-attainment - parameter estimates.*

| | β | S.E. | CI lower | CI upper | p |
|---|---|---|---|---|---|
| **Fixed Part** | | | | | |
| Intercept | 0.799 | 0.232 | 0.344 | 1.253 | ** |
| PrePost | 0.081 | 0.234 | -0.378 | 0.540 | |
| Medium intensity, low complexity | 0.121 | 0.259 | -0.387 | 0.629 | |
| High intensity, low complexity | -0.062 | 0.267 | -0.585 | 0.462 | |
| Low intensity, high complexity | -0.115 | 0.253 | -0.610 | 0.380 | |
| Medium intensity, high complexity | -0.081 | 0.265 | -0.601 | 0.439 | |
| High intensity, high complexity | -0.033 | 0.255 | -0.532 | 0.467 | |
| PrePost*Medium intensity, low complexity | 0.149 | 0.326 | -0.490 | 0.787 | |
| PrePost*High intensity, low complexity | 0.395 | 0.333 | -0.257 | 1.047 | |
| PrePost*Low intensity, high complexity | 0.174 | 0.321 | -0.454 | 0.803 | |
| PrePost*Medium intensity, high complexity | 0.274 | 0.330 | -0.373 | 0.921 | |
| PrePost*High intensity, high complexity | 0.601 | 0.324 | -0.034 | 1.236 | |
| Below ARE | -0.193 | 0.400 | -0.977 | 0.591 | |
| Above ARE | 0.491 | 0.281 | -0.060 | 1.041 | |
| PrePost*Below ARE | 0.403 | 0.521 | -0.617 | 1.424 | |
| PrePost*Above ARE | 0.076 | 0.344 | -0.599 | 0.751 | |
| Medium intensity, low complexity*Below ARE | 0.305 | 0.609 | -0.889 | 1.498 | |
| High intensity, low complexity*Below ARE | 0.312 | 0.550 | -0.766 | 1.391 | |
| Low intensity, high complexity*Below ARE | -0.220 | 0.541 | -1.280 | 0.841 | |
| Medium intensity, high complexity* Below ARE | 0.270 | 0.564 | -0.836 | 1.377 | |
| High intensity, high complexity*Below ARE | 0.188 | 0.521 | -0.833 | 1.208 | |
| Medium intensity, low complexity*Above ARE | 0.141 | 0.385 | -0.613 | 0.896 | |
| High intensity, low complexity*Above ARE | -0.117 | 0.393 | -0.887 | 0.653 | |
| Low intensity, high complexity*Above ARE | 0.334 | 0.376 | -0.403 | 1.070 | |
| Medium intensity, high complexity* Above ARE | 0.554 | 0.397 | -0.225 | 1.333 | |
| High intensity, high complexity*Above ARE | 0.064 | 0.378 | -0.676 | 0.805 | |
| PrePost*Medium intensity, low complexity* Below ARE | -0.180 | 0.776 | -1.700 | 1.341 | |
| PrePost*High intensity, low complexity*Below ARE | -0.579 | 0.713 | -1.977 | 0.819 | |
| PrePost*Low intensity, high complexity*BelowARE | -0.355 | 0.714 | -1.755 | 1.045 | |
| PrePost*Medium intensity, high complexity* Below ARE | -0.309 | 0.723 | -1.726 | 1.108 | |

| | β | S.E. | CI lower | CI upper | p |
|---|---|---|---|---|---|
| PrePost*High intensity, high complexity* Below ARE | -0.148 | 0.690 | -1.499 | 1.203 | |
| PrePost*Medium intensity, low complexity* Above ARE | -0.141 | 0.488 | -1.097 | 0.816 | |
| PrePost*High intensity, low complexity*Above ARE | -0.360 | 0.494 | -1.328 | 0.608 | |
| PrePost*Low intensity, high complexity*Above ARE | -0.663 | 0.477 | -1.597 | 0.272 | |
| PrePost*Medium intensity, high complexity* Above ARE | -0.867 | 0.495 | -1.838 | 0.103 | |
| PrePost*High intensity, high complexity* Above ARE | -0.358 | 0.484 | -1.306 | 0.590 | |
| **Random Part** | | | | | |
| Level: ID | | | | | |
| Intercept variance | 0.073 | 0.024 | | | ** |
| Level: Week | | | | | |
| Intercept variance | 0.810 | 0.073 | | | *** |
| Intercept/slope covariance | -0.823 | 0.082 | | | *** |
| Slope variance | 1.282 | 0.116 | | | *** |
| Units: ID | 73 | | | | |
| Units: Week | 394 | | | | |
| Units: Observations | 34,569 | | | | |

Significance values *=p<0.05, **=p<0.01, ***=p<0.001

$$
\begin{aligned}
On-task_{ijk} = {} & \beta_{0jk} + \beta_{1j}PrePost_{ijk} + \beta_2 Medium-Low_{jk} + \beta_3 High-Low_{jk} \\
& + \beta_4 Low-High_{jk} + \beta_5 Medium-High_{jk} + \beta_6 High-High_{jk} \\
& + \beta_7 PrePost * Medium-Low_{ijk} + \beta_8 PrePost * High-Low_{ijk} \\
& + \beta_9 PrePost * Low-High_{ijk} + \beta_{10} PrePost * Medium-High_{ijk} \\
& + \beta_{11} PrePost * High-High_{ijk} + \beta_{12} Below\ ARE_k + \beta_{13} Above\ ARE_k \\
& + \beta_{14} PrePost * Below\ ARE_{ijk} + \beta_{15} PrePost * Above\ ARE_{ijk} \\
& + \beta_{16} Medium-Low * Below\ ARE_{jk} + \beta_{17} High-Low * Below\ ARE_{jk} \\
& + \beta_{18} Low-High * Below\ ARE_{jk} + \beta_{19} Medium-High * Below\ ARE_{jk} \\
& + \beta_{20} High-High * Below\ ARE_{jk} + \beta_{21} Medium-Low * Above\ ARE_{jk} \\
& + \beta_{22} High-Low * Above\ ARE_{jk} + \beta_{23} Low-High * Above\ ARE_{jk} \\
& + \beta_{24} Medium-High * Above\ ARE_{jk} + \beta_{25} High-High * Above\ ARE_{jk} \\
& + \beta_{26} PrePost * Medium-Low * Below\ ARE_{ijk} \\
& + \beta_{27} PrePost * High-Low * Below\ ARE_{ijk} \\
& + \beta_{28} PrePost * Low-High * Below\ ARE_{ijk} \\
& + \beta_{29} PrePost * Medium-High * Below\ ARE_{ijk} \\
& + \beta_{30} PrePost * High-High * Below\ ARE_{ijk} \\
& + \beta_{31} PrePost * Medium-Low * Above\ ARE_{ijk} \\
& + \beta_{32} PrePost * High-Low * Above\ ARE_{ijk} \\
& + \beta_{33} PrePost * Low-High * Above\ ARE_{ijk} \\
& + \beta_{34} PrePost * Medium-High * Above\ ARE_{ijk} \\
& + \beta_{35} PrePost * High-High * Above\ ARE_{ijk} + \beta_{36-55} control\ variables
\end{aligned}
$$

$$\beta_{0jk} = \beta_0 + \nu_{0k} + \upsilon_{0jk}$$

$$\beta_{1j} = \beta_1 + \upsilon_{1PrePost_{jk}}$$

$$
\begin{aligned}
On-task_{ijk} = {} & \mathbf{0.799} + \mathbf{0.081}PrePost_{ijk} + \mathbf{0.121}Medium-Low_{jk} + \mathbf{-0.062}High-Low_{jk} \\
& + \mathbf{-0.115}Low-High_{jk} + \mathbf{-0.081}Medium-High_{jk} + \mathbf{-0.033}High-High_{jk} \\
& + \mathbf{0.149}PrePost*Medium-Low_{ijk} + \mathbf{0.395}PrePost*High-Low_{ijk} \\
& + \mathbf{0.174}PrePost*Low-High_{ijk} + \mathbf{0.274}PrePost*Medium-High_{ijk} \\
& + \mathbf{0.601}PrePost*High-High_{ijk} + \mathbf{-0.193}Below\ ARE_{k} + \mathbf{0.491}Above\ ARE \\
& + \mathbf{0.403}PrePost*Below\ ARE_{ijk} + \mathbf{0.076}PrePost*Above\ ARE_{ijk} \\
& + \mathbf{0.305}Medium-Low*Below\ ARE_{jk} + \mathbf{0.312}High-Low*Below\ ARE_{jk} \\
& + \mathbf{-0.220}Low-High*Below\ ARE_{jk} + \mathbf{0.270}Medium-High*Below\ ARE_{jk} \\
& + \mathbf{0.188}High-High*Below\ ARE_{jk} + \mathbf{0.141}Medium-Low*Above\ ARE_{jk} \\
& + \mathbf{-0.117}High-Low*Above\ ARE_{jk} + \mathbf{0.334}Low-High*Above\ ARE_{jk} \\
& + \mathbf{0.554}Medium-High*Above\ ARE_{jk} + \mathbf{0.064}High-High*Above\ ARE_{jk} \\
& + \mathbf{-0.180}PrePost*Medium-Low*Below\ ARE_{ijk} \\
& + \mathbf{-0.579}PrePost*High-Low*Below\ ARE_{ijk} \\
& + \mathbf{-0.355}PrePost*Low-High*Below\ ARE_{ijk} \\
& + \mathbf{-0.309}PrePost*Medium-High*Below\ ARE_{ijk} \\
& + \mathbf{-0.148}PrePost*High-High*Below\ ARE_{ijk} \\
& + \mathbf{-0.141}PrePost*Medium-Low*Above\ ARE_{ijk} \\
& + \mathbf{-0.360}PrePost*High-Low*Above\ ARE_{ijk} \\
& + \mathbf{-0.663}PrePost*Low-High*Above\ ARE_{ijk} \\
& + \mathbf{-0.867}PrePost*Medium-High*Above\ ARE_{ijk} \\
& + \mathbf{-0.358}PrePost*High-High*Above\ ARE_{ijk} + \beta_{36-55}control\ variables \\
& + \upsilon_{1PrePost_{jk}} + \nu_{0k} + \upsilon_{0jk}
\end{aligned}
$$

$$[\nu_{0k}] \sim N(0,\Omega): \Omega_{\nu} = [\sigma^2_{\nu 0}] = 0.073$$

$$\begin{bmatrix}\upsilon_{0jk}\\ \upsilon_{1jk}\end{bmatrix} \sim N(0,\Omega_{\upsilon}): \Omega_{\upsilon} = \begin{bmatrix}\sigma^2_{\upsilon 0} & \\ \sigma_{\upsilon 01} & \sigma^2_{\upsilon 1}\end{bmatrix} = \begin{bmatrix}0.810 & \\ \mathbf{-0.823} & 1.282\end{bmatrix}$$

*Equation 35. Model 5.2. Intensity-by-complexity-by-attainment.*

Although only one or two conditions for each group showed statistically significant effects, small to medium effects were identified for four intervention conditions for participants in the 'below ARE' group, three in the 'at ARE' group, and two in the 'above ARE' group (Figure 33). The only negative effect identified was a small effect for participants in the 'above ARE' category; they displayed a lower proportion of on-task behaviour after taking part in the high intensity, low complexity intervention

lesson (d=-0.24, OR=0.65). One medium effect size was found for participants who scored 'below ARE' after high intensity, high complexity lessons (d=0.52, OR=2.55). Contrary to the hypothesis, those with lower levels of attainment benefited more from participation in PA in terms of the number of intervention lessons which had a positive effect on subsequent on-task behaviour. Although the difference in effect varied across conditions, it seems that high achievers experienced consistently smaller effects after high complexity lessons.

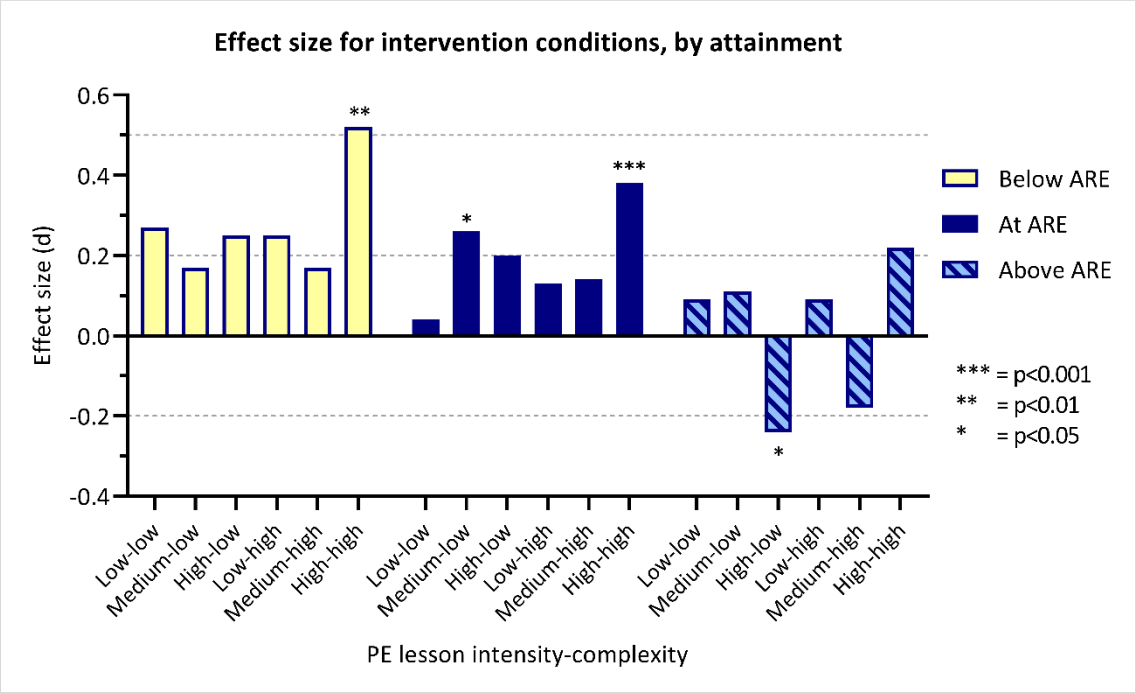



*Figure 33. Effect of PE lesson intensity-by-complexity-by-attainment.*

*Weekly sports participation duration.* For both groups (more or less active than the sample mean) the only PE lesson after which on-task behaviour improved significantly was the high intensity, high complexity condition ($\chi^2_{<\ active\ than\ sample\ mean}$=12.242, p<0.001; $\chi^2_{\geq active\ than\ sample\ mean}$=5.715, p<0.05). Parameter estimates for the predictor variables in model 5.3 can be found in Table 29 and Equation 36. Full parameter estimates are included in appendix O.

*Table 29. Model 5.3: Intensity-by-complexity-by-sports participation - parameter estimates.*

| | β | S.E. | CI lower | CI upper | p |
|---|---|---|---|---|---|
| **Fixed Part** | | | | | |
| Intercept | 0.655 | 0.211 | 0.241 | 1.069 | ** |
| PrePost | 0.276 | 0.208 | -0.133 | 0.685 | |
| Medium intensity, low complexity | -0.078 | 0.236 | -0.541 | 0.384 | |
| High intensity, low complexity | 0.159 | 0.236 | -0.303 | 0.622 | |
| Low intensity, high complexity | 0.445 | 0.234 | -0.013 | 0.903 | |
| Medium intensity, high complexity | 0.101 | 0.221 | -0.333 | 0.535 | |
| High intensity, high complexity | -0.003 | 0.224 | -0.441 | 0.436 | |
| PrePost*Medium intensity, low complexity | 0.033 | 0.302 | -0.559 | 0.625 | |
| PrePost*High intensity, low complexity | -0.005 | 0.303 | -0.599 | 0.590 | |
| PrePost*Low intensity, high complexity | -0.065 | 0.302 | -0.657 | 0.527 | |
| PrePost*Medium intensity, high complexity | -0.058 | 0.289 | -0.624 | 0.509 | |
| PrePost*High intensity, high complexity | 0.377 | 0.294 | -0.200 | 0.954 | |
| Sports participation duration (min) > sample mean | 0.050 | 0.266 | -0.471 | 0.571 | |
| PrePost*Sports participation duration (min) > sample mean | -0.311 | 0.336 | -0.971 | 0.348 | |
| Medium intensity, low complexity* Sports participation duration (min) > sample mean | -0.001 | 0.366 | -0.719 | 0.717 | |
| High intensity, low complexity* Sports participation duration (min) > sample mean | -0.076 | 0.371 | -0.803 | 0.651 | |
| Low intensity, high complexity* Sports participation duration (min) > sample mean | -0.567 | 0.365 | -1.283 | 0.150 | |
| Medium intensity, high complexity* Sports participation duration (min) > sample mean | -0.371 | 0.354 | -1.065 | 0.324 | |
| High intensity, high complexity* Sports participation duration (min) > sample mean | 0.046 | 0.352 | -0.644 | 0.736 | |
| PrePost*Medium intensity, low complexity* Sports participation duration (min) > sample mean | 0.398 | 0.474 | -0.531 | 1.328 | |
| PrePost*High intensity, low complexity* Sports participation duration (min) > sample mean | -0.104 | 0.476 | -1.036 | 0.828 | |
| PrePost*Low intensity, high complexity* Sports participation duration (min) > sample mean | 0.320 | 0.475 | -0.610 | 1.251 | |
| PrePost*Medium intensity, high complexity* Sports participation duration (min) > sample mean | -0.062 | 0.464 | -0.972 | 0.848 | |
| PrePost*High intensity, high complexity* Sports participation duration (min) > sample mean | 0.165 | 0.463 | -0.744 | 1.073 | |

| | β | S.E. | CI lower | CI upper | p |
|---|---|---|---|---|---|
| **Random Part** | | | | | |
| Level: ID | | | | | |
| Intercept variance | 0.067 | 0.023 | | | ** |
| Level: Week | | | | | |
| Intercept variance | 0.797 | 0.072 | | | *** |
| Intercept/slope covariance | -0.843 | 0.083 | | | *** |
| Slope variance | 1.339 | 0.120 | | | *** |
| Units: ID | 73 | | | | |
| Units: Week | 394 | | | | |
| Units: Observations | 34,569 | | | | |

Significance values *=p<0.05, **=p<0.01, ***=p<0.001

$$
\begin{aligned}
On-task_{ijk} = {} & \beta_{0jk} + \beta_{1j} PrePost_{ijk} + \beta_2 Medium-Low_{jk} + \beta_3 High-Low_{jk} \\
& + \beta_4 Low-High_{jk} + \beta_5 Medium-High_{jk} + \beta_6 High-High_{jk} \\
& + \beta_7 PrePost * Medium-Low_{ijk} + \beta_8 PrePost * High-Low_{ijk} \\
& + \beta_9 PrePost * Low-High_{ijk} + \beta_{10} PrePost * Medium-High_{ijk} \\
& + \beta_{11} PrePost * High-High_{ijk} \\
& + \beta_{12} Sports\ participation\ duration > \ sample\ mean_k \\
& + \beta_{13} PrePost * Sports\ participation\ duration > \ sample\ mean_{ijk} \\
& + \beta_{14} Medium-Low * Sports\ participation\ duration > \ sample\ mean_{jk} \\
& + \beta_{15} High-Low * Sports\ participation\ duration > \ sample\ mean_{jk} \\
& + \beta_{16} Low-High * Sports\ participation\ duration > \ sample\ mean_{jk} \\
& + \beta_{17} Medium-High * Sports\ participation\ duration > \ sample\ mean_{jk} \\
& + \beta_{18} High-High * Sports\ participation\ duration > \ sample\ mean_{jk} \\
& + \beta_{19} PrePost * Medium-Low * Sports\ participation\ duration \\
& > \ sample\ mean_{ijk} \\
& + \beta_{20} PrePost * High-Low * Sports\ participation\ duration \\
& > \ sample\ mean_{ijk} \\
& + \beta_{21} PrePost * Low-High * Sports\ participation\ duration \\
& > \ sample\ mean_{ijk} \\
& + \beta_{22} PrePost * Medium-High * Sports\ participation\ duration \\
& > \ sample\ mean_{ijk} \\
& + \beta_{23} PrePost * High-High * Sports\ participation\ duration \\
& > \ sample\ mean_{ijk} \ + \beta_{24-46} control\ variables
\end{aligned}
$$

$$\beta_{0jk} = \beta_0 + \nu_{0k} + \upsilon_{0jk}$$

$$\beta_{1j} = \beta_1 + \upsilon_{1PrePost_{jk}}$$

$$\begin{aligned} On-task_{ijk} = {} & \mathbf{0.655} + \mathbf{0.276}PrePost_{ijk} + \mathbf{-0.078}Medium-Low_{jk} + \mathbf{0.159}High-Low_{jk} \\ & + \mathbf{0.445}Low-High_{jk} + \mathbf{0.101}Medium-High_{jk} + \mathbf{-0.003}High-High_{jk} \\ & + \mathbf{0.033}PrePost*Medium-Low_{ijk} + \mathbf{-0.005}PrePost*High-Low_{ijk} \\ & + \mathbf{-0.065}PrePost*Low-High_{ijk} + \mathbf{-0.058}PrePost*Medium-High_{ijk} \\ & + \mathbf{0.377}PrePost*High-High_{ijk} \\ & + \mathbf{0.050}Sports\ participation\ duration > sample\ mean_k \\ & + \mathbf{-0.311}PrePost*Sports\ participation\ duration > sample\ mean_{ijk} \\ & + \mathbf{-0.001}Medium-Low*Sports\ participation\ duration > sample\ mean_{jk} \\ & + \mathbf{-0.076}High-Low*Sports\ participation\ duration > sample\ mean_{jk} \\ & + \mathbf{-0.567}Low-High*Sports\ participation\ duration > sample\ mean_{jk} \\ & + \mathbf{-0.371}Medium-High*Sports\ participation\ duration > sample\ mean_{jk} \\ & + \mathbf{0.046}High-High*Sports\ participation\ duration > sample\ mean_{jk} \\ & + \mathbf{0.398}PrePost*Medium-Low*Sports\ participation\ duration \\ & > sample\ mean_{ijk} \\ & + \mathbf{-0.104}PrePost*High-Low*Sports\ participation\ duration \\ & > sample\ mean_{ijk} \\ & + \mathbf{0.320}PrePost*Low-High*Sports\ participation\ duration \\ & > sample\ mean_{ijk} \\ & + \mathbf{-0.062}PrePost*Medium-High*Sports\ participation\ duration \\ & > sample\ mean_{ijk} \\ & + \mathbf{0.165}PrePost*High-High*Sports\ participation\ duration \\ & > sample\ mean_{ijk} + \beta_{24-46}control\ variables + \upsilon_{1PrePost_{jk}} + \nu_{0k} + \upsilon_{0jk} \end{aligned}$$

$$[\nu_{0k}] \sim N(0, \Omega): \Omega_\nu = [\sigma^2_{\nu 0}] = 0.067$$

$$\begin{bmatrix} \upsilon_{0jk} \\ \upsilon_{1jk} \end{bmatrix} \sim N(0, \Omega_\upsilon): \Omega_\upsilon = \begin{bmatrix} \sigma^2_{\upsilon 0} & \\ \sigma_{\upsilon 01} & \sigma^2_{\upsilon 1} \end{bmatrix} = \begin{bmatrix} 0.797 & \\ \mathbf{-0.843} & 1.339 \end{bmatrix}$$

*Equation 36. Model 5.3. Intensity-by-complexity-by-sports participation.*

Effect sizes differed somewhat between those who participated in sports for longer or shorter than the sample mean (Figure 34). However, there was only one condition where one group and not the other spent a greater proportion of the subsequent lesson on-task. There was a small effect (d=0.22,

OR=1.49) for the more active group after participation in the medium intensity, low complexity lesson, and no effect for the less active group (d=0.17, OR=1.36). For the high intensity, high complexity lesson, a larger effect size was found for the less active group (d=0.36, OR=1.93) compared to the more active group (d=0.28, OR=1.66), but the effect size for both groups was small. The hypothesis that those who are less active benefit more from participation in PA was rejected.

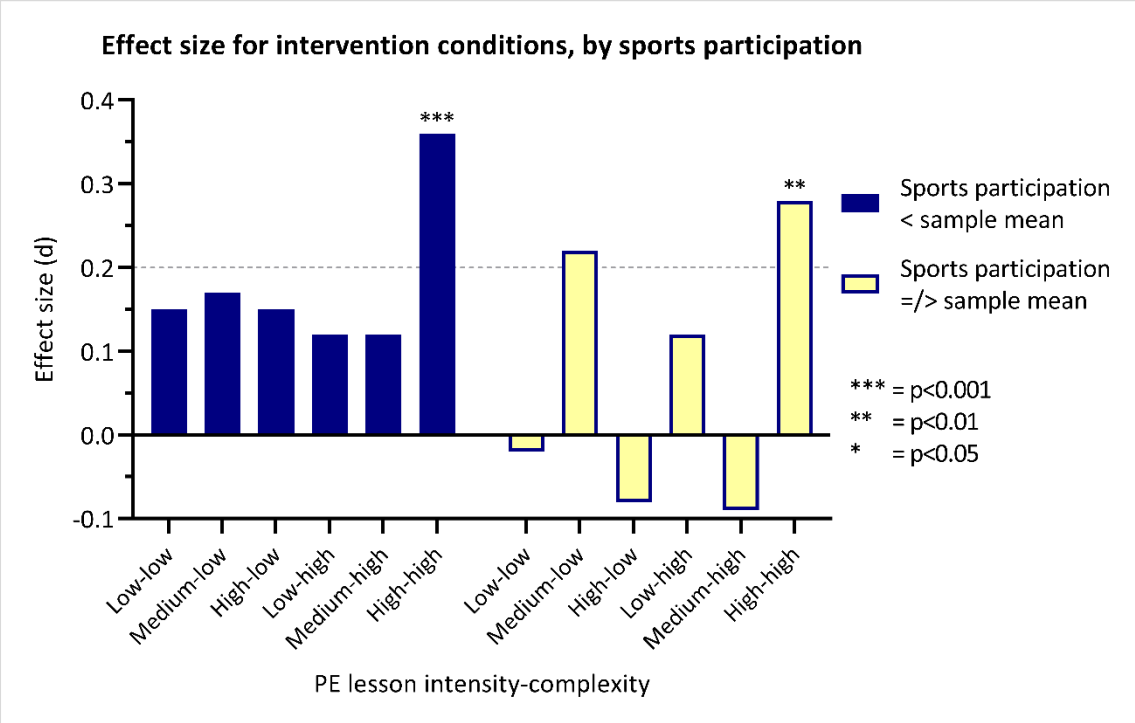



*Figure 34. Effect of PE lesson intensity-by-complexity-by-sports participation.*

*Gender*. Whereas there were no conditions where boys were significantly more on-task after PE than before, girls were more on-task after medium intensity-low complexity lessons ($\chi^2$=4.728, p<0.05), and high intensity-high complexity lessons ($\chi^2$=17.640, p<0.001). Parameter estimates for the predictor variables in model 5.4 can be found in Table 30 and Equation 37. Full parameter estimates are included in appendix P.

*Table 30. Model 5.4: Intensity-by-complexity-by-gender- parameter estimates.*

| | | β | S.E. | CI lower | CI upper | p |
|---|---|---|---|---|---|---|
| **Fixed Part** | | | | | | |
| | Intercept | 1.064 | 0.242 | 0.590 | 1.539 | *** |
| | PrePost | -0.163 | 0.256 | -0.665 | 0.339 | |
| | Medium intensity, low complexity | -0.257 | 0.299 | -0.842 | 0.329 | |
| | High intensity, low complexity | -0.036 | 0.296 | -0.617 | 0.544 | |
| | Low intensity, high complexity | -0.128 | 0.292 | -0.700 | 0.444 | |
| | Medium intensity, high complexity | -0.269 | 0.276 | -0.810 | 0.273 | |
| | High intensity, high complexity | -0.036 | 0.279 | -0.583 | 0.510 | |
| | PrePost*Medium intensity, low complexity | 0.376 | 0.374 | -0.356 | 1.108 | |
| | PrePost*High intensity, low complexity | -0.098 | 0.369 | -0.821 | 0.625 | |
| | PrePost*Low intensity, high complexity | 0.290 | 0.369 | -0.432 | 1.013 | |
| | PrePost*Medium intensity, high complexity | 0.120 | 0.353 | -0.573 | 0.813 | |
| | PrePost*High intensity, high complexity | 0.472 | 0.359 | -0.231 | 1.176 | |
| | Female | -0.179 | 0.268 | -0.703 | 0.345 | |
| | PrePost*Female | 0.517 | 0.332 | -0.134 | 1.168 | |
| | Medium intensity, low complexity*Female | 0.306 | 0.377 | -0.432 | 1.044 | |
| | High intensity, low complexity*Female | 0.285 | 0.377 | -0.455 | 1.024 | |
| | Low intensity, high complexity*Female | 0.534 | 0.372 | -0.194 | 1.263 | |
| | Medium intensity, high complexity*Female | 0.359 | 0.356 | -0.339 | 1.056 | |
| | High intensity, high complexity*Female | 0.095 | 0.357 | -0.605 | 0.794 | |
| | PrePost*Medium intensity, low complexity*Female | -0.332 | 0.477 | -1.267 | 0.602 | |
| | PrePost*High intensity, low complexity*Female | 0.076 | 0.475 | -0.856 | 1.007 | |
| | PrePost*Low intensity, high complexity*Female | -0.350 | 0.474 | -1.279 | 0.580 | |
| | PrePost*Medium intensity, high complexity*Female | -0.333 | 0.458 | -1.232 | 0.565 | |
| | PrePost*High intensity, high complexity*Female | -0.054 | 0.462 | -0.960 | 0.852 | |
| **Random Part** | Level: ID | | | | | |
| | Intercept variance | 0.064 | 0.023 | | | ** |
| | Level: Week | | | | | |
| | Intercept variance | 0.812 | 0.073 | | | *** |
| | Intercept/slope covariance | -0.829 | 0.083 | | | *** |
| | Slope variance | 1.324 | 0.119 | | | *** |
| | Units: ID | 73 | | | | |
| | Units: Week | 394 | | | | |
| | Units: Observations | 34,569 | | | | |

Significance values *=p<0.05, **=p<0.01, ***=p<0.001

$$On-task_{ijk} = \beta_{0jk} + \beta_{1j}PrePost_{ijk} + \beta_2 Medium-Low_{jk} + \beta_3 High-Low_{jk}$$
$$+ \beta_4 Low-High_{jk} + \beta_5 Medium-High_{jk} + \beta_6 High-High_{jk}$$
$$+ \beta_7 PrePost * Medium-Low_{ijk} + \beta_8 PrePost * High-Low_{ijk}$$
$$+ \beta_9 PrePost * Low-High_{ijk} + \beta_{10} PrePost * Medium-High_{ijk}$$
$$+ \beta_{11} PrePost * High-High_{ijk} + \beta_{12} Female_k + \beta_{13} PrePost * Female_{ijk}$$
$$+ \beta_{14} Medium-Low * Female_{jk} + \beta_{15} High-Low * Female_{jk}$$
$$+ \beta_{16} Low-High * Female_{jk} + \beta_{17} Medium-High * Female_{jk}$$
$$+ \beta_{18} High-High * Female_{jk} + \beta_{19} PrePost * Medium-Low * Female_{ijk}$$
$$+ \beta_{20} PrePost * High-Low * Female_{ijk}$$
$$+ \beta_{21} PrePost * Low-High * Female_{ijk}$$
$$+ \beta_{22} PrePost * Medium-High * Female_{ijk}$$
$$+ \beta_{23} PrePost * High-High * Female_{ijk} + \beta_{24-46} control\ variables$$

$$\beta_{0jk} = \beta_0 + \nu_{0k} + \upsilon_{0jk}$$

$$\beta_{1j} = \beta_1 + \upsilon_{1PrePost_{jk}}$$

$$On-task_{ijk} = \mathbf{1.064} + \mathbf{-0.163} PrePost_{ijk} + \mathbf{-0.257} Medium-Low_{jk}$$
$$+ \mathbf{-0.036} High-Low_{jk} + \mathbf{-0.128} Low-High_{jk} + \mathbf{-0.269} Medium-High_{jk}$$
$$+ \mathbf{-0.036} High-High_{jk} + \mathbf{0.376} PrePost * Medium-Low_{ijk}$$
$$+ \mathbf{-0.098} PrePost * High-Low_{ijk} + \mathbf{0.290} PrePost * Low-High_{ijk}$$
$$+ \mathbf{0.120} PrePost * Medium-High_{ijk} + \mathbf{0.472} PrePost * High-High_{ijk}$$
$$+ \mathbf{-0.179} Female_k + \mathbf{0.517} PrePost * Female_{ijk}$$
$$+ \mathbf{0.306} Medium-Low * Female_{jk} + \mathbf{0.285} High-Low * Female_{jk}$$
$$+ \mathbf{0.534} Low-High * Female_{jk} + \mathbf{0.359} Medium-High * Female_{jk}$$
$$+ \mathbf{0.095} High-High * Female_{jk}$$
$$+ \mathbf{-0.332} PrePost * Medium-Low * Female_{ijk}$$
$$+ \mathbf{0.076} PrePost * High-Low * Female_{ijk}$$
$$+ \mathbf{-0.350} PrePost * Low-High * Female_{ijk}$$
$$+ \mathbf{-0.333} PrePost * Medium-High * Female_{ijk}$$
$$+ \mathbf{-0.054} PrePost * High-High * Female_{ijk} + \beta_{24-47} control\ variables$$
$$+ \upsilon_{1PrePost_{jk}} + \nu_{0k} + \upsilon_{0jk}$$

$$[\nu_{0k}] \sim N(0, \Omega): \Omega_\nu = [\sigma^2_{\nu 0}] = 0.064$$

$$\begin{bmatrix} \upsilon_{0jk} \\ \upsilon_{1jk} \end{bmatrix} \sim N(0, \Omega_\upsilon): \Omega_\upsilon = \begin{bmatrix} \sigma^2_{\upsilon 0} & \\ \sigma_{\upsilon 01} & \sigma^2_{\upsilon 1} \end{bmatrix} = \begin{bmatrix} 0.812 & \\ \mathbf{-0.829} & 1.324 \end{bmatrix}$$

*Equation 37. Model 5.4. Intensity-by-complexity-by-gender.*

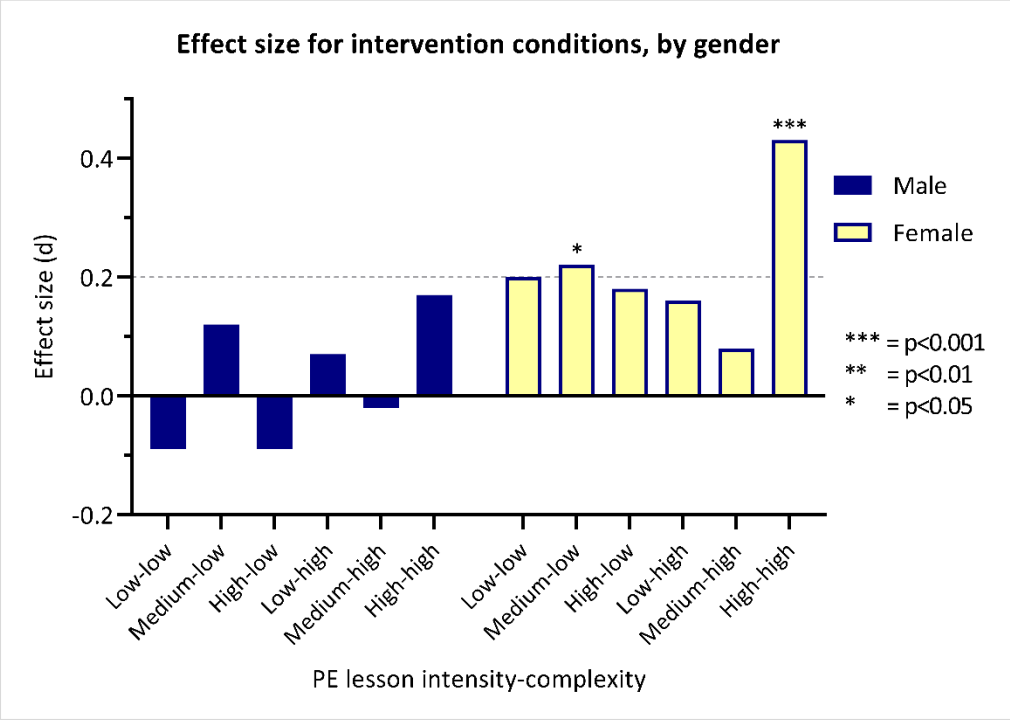



*Figure 35. Effect of PE lesson intensity-by-complexity-by-gender.*

Effect sizes differed between boys and girls (Figure 35), with no effect sizes for boys reaching the d=0.2 cut-point (d's=-0.14 to 0.17). Girls on the other hand benefited from three of the intervention conditions with effect sizes reaching or exceeding d=0.2 (low intensity, low complexity: d=0.2, OR=1.43; medium intensity, low complexity: d=0.22, OR=1.49; high intensity, high complexity: d=0.43, OR=2.16). Low intensity, low complexity had an effect size of d=0.2 despite not reaching statistical significance. The lack of statistical significance may be due to the small size of the sub-sample. The hypothesis that boys would benefit most from taking part in PA was rejected, as girls benefited more.

To summarise, this section of the results chapter has provided answers to research questions one to four. The hypotheses for questions one and two were rejected, as no main effects of intensity or

complexity were found. The hypothesis for question three was accepted, as there was an interaction effect of intensity-by-complexity, but the hypothesised direction of the dose-response relationship was not established. Research question four presented a complex situation. Individual differences existed for gender, attainment, and pre-PE level of on-task behaviour, but not for weekly sports participation duration. However, the nature of the interaction between individual characteristics and PA was opposite to the hypotheses for gender and attainment, with girls and those with the lowest achievement levels profiting the most from PA participation. The hypothesis that those with the lowest levels of pre-PE on-task behaviour would benefit greatest was accepted. Research question five, regarding mediation of the relationship between PA and on-task behaviour by emotion states, will be answered in the next section.

## 4.4 Physical Activity, Emotion States, and Behavioural Engagement

In this section, results are provided of the analyses of emotion data derived from the repeated self-reports completed by participants. First, descriptives of the emotion scores are given. Then, the step-by-step building of the SEM from factor analysis to mediation models is presented.

### 4.4.1 Emotion State Descriptives

*Table 31. Percentage high emotion ratings in classroom and PE reports*

| | | Classroom | PE |
|---|---|---|---|
| Rated 3 stars or more out of 5 | Excited | 68.0 | 65.9 |
| | Happy | 79.9 | 82.0 |
| | Calm | 76.2 | 68.7 |
| | Relaxed | 75.9 | 68.2 |
| | Bored | 38.4 | 30.1 |
| | Sad | 17.2 | 17.3 |
| | Angry | 15.6 | 17.2 |
| | Nervous | 17.9 | 16.9 |

Table 31 displays the percentages of emotions reported at 3 or more stars out of 5. Fewer than 40% of questionnaires report high levels of negative emotions, and over 65% of questionnaires reports high levels of positive emotions. Correspondingly, the negative emotions items were negatively skewed, and all positive emotions were positively skewed.

A large number of self-reports had missing data, where participants had missed one or more questions. Some reports were submitted with the same rating for each of the 8 emotions, in which case it was deemed that the participant had misunderstood the questions or had rushed to complete the questionnaire without reading the questions, and the data was recoded as missing. For missingness of each emotion variable see Table 32.

*Table 32. Missing data in self-reports.*

| Self-report item | Number missing classroom reports (% of total classroom reports) | Number missing PE reports (% of total PE reports) |
|---|---|---|
| PE enjoyment | N/A | 8 (1.4%) |
| Excited | 322 (13.7%) | 30 (5.2%) |
| Happy | 325 (13.9%) | 31 (5.4%) |
| Calm | 334 (14.3%) | 33 (5.7%) |
| Relaxed | 345 (14.7%) | 39 (6.8%) |
| Bored | 352 (15.0%) | 57 (9.9%) |
| Sad | 454 (19.4%) | 74 (12.8%) |
| Angry | 461 (19.7%) | 70 (12.2%) |
| Nervous | 417 (17.8%) | 67 (11.6%) |

### 4.4.2 Mediation Models of Physical Activity, Emotion State, and Behavioural Engagement

Exploratory and confirmatory factor analyses were conducted first, then questionnaire level aggregated models were specified. These were further developed into lesson level aggregated models with PE lessons as predictors of emotions, and finally into mediation models with PE lesson variables as predictors, emotion scores as mediators, and on-task behaviour as the DV.

*Table 33. Overview of SEM goodness of fit.*

| | $\chi^2$ (p) | AIC | BIC | RMSEA | CFI | SRMR | | |
|---|---|---|---|---|---|---|---|---|
| | | | | | | Within (Weeks) | Between (Weeks) | Between (ID) |
| EFA 8 emotions, 3 factors | 172.93 (***) | 65529.74 | 65750.58 | 0.091 | 0.980 | | | |
| EFA 7 emotions, 3 factors (excited excluded) | 37.89 (***) | 55472.62 | 55662.47 | 0.065 | 0.996 | | | |
| EFA 7 emotions, 3 factors (bored excluded) | 17.68 (***) | 58591.70 | 58782.67 | 0.041 | 0.997 | | | |
| EFA 6 emotions, 2 factors | 30.15 (***) | 48673.71 | 48810.09 | 0.049 | 0.995 | | | |
| CFA 6 emotions, 2 factors – 3-level structure | 105.04 (***) | 42982.97 | 43249.81 | 0.035 | 0.950 | 0.055 | 0.060 | 0.111 |
| Invariance model: | | | | | | | | |
| positive emotions | 521.95 (***) | 26525.08 | 26910.48 | 0.064 | 0.951 | - | 0.062 | 0.000 |
| negative emotions | 808.75 (***) | 21380.49 | 21764.53 | 0.087 | 0.912 | - | 0.074 | 0.000 |
| Correlation (questionnaire level) | 176.88 (***) | 14035.69 | 14363.97 | 0.078 | 0.968 | - | 0.100 | 0.000 |
| Autoregression: | | | | | | | | |
| classroom only | 362.74 (***) | 11981.21 | 12151.36 | 0.110 | 0.912 | - | 0.100 | 0.000 |
| classroom and PE | 453.70 (***) | 14256.50 | 14309.87 | 0.102 | 0.908 | - | 0.100 | 0.000 |
| Correlation (lesson level) | 328.27 (***) | 19809.49 | 20139.23 | 0.060 | 0.959 | - | 0.047 | 0.000 |
| Autoregression: | | | | | | | | |
| with mediation paths | 410.88 (***) | 19864.10 | 20133.89 | 0.065 | 0.947 | - | 0.078 | 0.000 |
| with cross-lags and mediation paths | 359.79 (***) | 19821.01 | 20107.92 | 0.060 | 0.955 | - | 0.051 | 0.000 |
| Autoregression with predictors: | | | | | | | | |
| PE conditions | 433.17 (***) | 19824.50 | 20154.24 | 0.046 | 0.957 | - | 0.046 | 0.000 |
| PE enjoyment | 542.84 (***) | 21529.66 | 21898.90 | 0.052 | 0.941 | - | 0.052 | 0.000 |
| Regression with behaviour outcomes | 598.05 (***) | 18558.95 | 18660.61 | 0.051 | 0.937 | - | 0.052 | 0.000 |

An overview of the stepwise development of the SEM and the goodness of fit indices for each model are displayed in Table 33. The first goodness of fit index reported is the $\chi^2$ statistic and its significance. A significant $\chi^2$ rejects the null hypothesis that the model fits perfectly to the population (Geiser, 2013). Next are the Information Criteria, the AIC and BIC. These can be used to allow comparison between models (ibid.). The relative goodness of fit indices must be used with caution, as not all models can be directly compared to each other. If a model is extended by the addition or removal of a parameter or regression/correlation path, the relative goodness of fit indices indicate if the model adaptation has improved the model's goodness of fit (shown by a lower value of the index), or not (ibid.). Finally, the Root Mean Square Error of Approximation (RMSEA), Comparative Fit Index (CFI), and Standardised Root Mean Square Residual (SRMR) are reported for approximate model fit (ibid.). On both the RMSEA and SRMR values below 0.05 are usually accepted as an indicator of good model fit, whereas the CFI has an approximate cut-off of 0.95 (ibid.). However, for small sample sizes these indices can be harsh, especially where models are complex, and must thus be interpreted with flexibility and in conjunction with the application of subjective judgement (Browne and Cudeck, 1992).

### Factor Analyses of Self-Report Ratings

Exploratory factor analysis of the self-reported emotion ratings was carried out in Mplus8 (Muthen and Muthen, 2017). Initially, scores for all eight reported emotions were used, specifying one-, two-, and three-factor options. The hypothesized four-factor solution did not converge due to exceeding the possible number of iterations, and model fit was best for the three-factor model (Table 34).

*Table 34. Goodness of fit indices for 8-emotion EFA.*

| | $\chi^2$ (p) | AIC | BIC | RMSEA | CFI |
|---|---|---|---|---|---|
| 1-factor | 3405.083 (***) | 68735.894 | 68879.142 | 0.242 | 0.593 |
| 2-factor | 1240.067 (***) | 66584.879 | 66769.907 | 0.181 | 0.852 |
| 3-factor | 172.930 (***) | 65529.742 | 65750.583 | 0.091 | 0.980 |

Significance values *=p<0.05, **=p<0.01, ***=p<0.001

A cut-off of 0.4 was applied throughout for factor loadings, as this is considered a substantial loading (Field, 2013). Factor loadings greater than 0.4 for the three-factor solution are displayed in Table 35. For all SEM results, standardised parameters are reported.

*Table 35. Standardised factor loadings for 3-factor model with 8 emotions.*

| | Factor 1 | Factor 2 | Factor 3 |
|---|---|---|---|
| Excited | | | |
| Happy | 0.707 | | |
| Calm | 0.623 | | |
| Relaxed | | 0.914 | |
| Bored | | 0.854 | |
| Sad | | | 0.435 |
| Angry | | | 0.799 |
| Nervous | | | 0.906 |

As 'excited' did not load onto any factors above 0.4, and model fit of the three-factor model was not perfect, the EFA was repeated on seven emotions. The seven-emotion EFA model fit indices can be found in Table 36.

*Table 36. Goodness of fit indices for 7-emotion EFA excluding 'excited'.*

| | $\chi^2$ (p) | AIC | BIC | RMSEA | CFI |
|---|---|---|---|---|---|
| 1-factor | 3080.415 (***) | 58493.136 | 58617.731 | 0.280 | 0.614 |
| 2-factor | 1006.528 (***) | 56431.249 | 56591.442 | 0.212 | 0.874 |
| 3-factor | 37.899 (***) | 55472.620 | 55662.479 | 0.065 | 0.996 |

Significance values *=p<0.05, **=p<0.01, ***=p<0.001

Again, the three-factor model achieved the best fit, though the loadings of the three-factor model did not make qualitative sense. The items 'calm' and 'relaxed' were split up from one another, and the item 'bored' seemed more strongly related to the positive emotions than the negative ones. The loadings in the two-factor model suggested emotions linked to a 'positive emotions' and a 'negative emotions' factor. However, this model did not have a good model fit, possibly due to the unexplained loading of the item 'bored' on the positive factor. To examine this, the seven-emotion EFA was

repeated with ‘excited’ and without ‘bored’. Factor loadings greater than 0.4 are displayed in Table 37.

*Table 37. Standardised factor loadings for 2- and 3-factor model with 7 emotions excluding ‘excited’.*

| | Factor 1 | Factor 2 | Factor 1 | Factor 2 | Factor 3 |
|---|---|---|---|---|---|
| Happy | 0.550 | | 0.776 | | |
| Calm | 0.677 | | 0.747 | | |
| Relaxed | 0.873 | | | 0.940 | |
| Bored | 0.867 | | | 0.801 | |
| Sad | | 0.435 | | | 0.414 |
| Angry | | 0.786 | | | 0.761 |
| Nervous | | 0.930 | | | 0.945 |

The three-factor model with excited included instead of bored achieved a good model fit on all indices (Table 38).

*Table 38. Goodness of fit indices for 7-emotion EFA, excluding ‘bored’.*

| | $\chi^2$ (p) | AIC | BIC | RMSEA | CFI |
|---|---|---|---|---|---|
| 1-factor | 2427.402 (***) | 60979.427 | 61104.747 | 0.244 | 0.576 |
| 2-factor | 233.978 (***) | 58798.002 | 58959.128 | 0.099 | 0.960 |
| 3-factor | 17.685 (***) | 58591.709 | 58782.673 | 0.041 | 0.997 |

Significance values *=p<0.05, **=p<0.01, ***=p<0.001

Still, when excited was included instead of bored, the factor loadings presented problems for interpretation, as can be seen in Table 39. The item ‘sad’ did not load onto any factor above the 0.4 cut-off applied, and ‘excited’ was the only item to load onto the second factor. Also, the loading for ‘nervous’ was greater than 1, whereas loadings normally range between -1 and 1 (Field, 2013). The next best solution was the two-factor model, which achieved a good fit only on the CFI index (>0.95). In this model, ‘sad’ loaded onto the same factor as ‘angry’ and ‘nervous’ to give a factor for negative emotions. ‘Excited’ did not load onto either factor.

*Table 39. Standardised factor loadings for 2- and 3-factor model with 7 emotions excluding 'bored'.*

| | Factor 1 | Factor 2 | Factor 1 | Factor 2 | Factor 3 |
|---|---|---|---|---|---|
| Happy | 0.717 | | 0.688 | | |
| Excited | | | | 0.767 | |
| Calm | 0.939 | | 0.952 | | |
| Relaxed | 0.580 | | 0.592 | | |
| Sad | | 0.420 | | | |
| Angry | | 0.791 | | | 0.717 |
| Nervous | | 0.919 | | | 1.010 |

A final EFA was carried out for six emotions; excluding both 'excited' and 'bored', with options for one-, two-, and three-factor models. Both the two- and three-factor solutions achieved good model fit on all indices. The three-factor solution achieved the best model fit, with the exception of the BIC, which was slightly better for the two-factor solution (Table 40).

*Table 40. Goodness of fit indices for 6-emotion EFA.*

| | $\chi^2$ (p) | AIC | BIC | RMSEA | CFI |
|---|---|---|---|---|---|
| 1-factor | 2179.773 (***) | 50813.327 | 50920.064 | 0.295 | 0.591 |
| 2-factor | 30.157 (***) | 48673.711 | 48810.097 | 0.049 | 0.995 |
| 3-factor | 3.868 (***) | 48655.421 | 48815.527 | 0.000 | 0.999 |

*Table 41. Standardised factor loadings for 2- and 3-factor model with 6 emotions.*

| | Factor 1 | Factor 2 | Factor 1 | Factor 2 | Factor 3 |
|---|---|---|---|---|---|
| Happy | 0.705 | | 0.610 | | |
| Calm | 0.952 | | 0.953 | | |
| Relaxed | 0.577 | | 0.591 | | |
| Sad | | 0.404 | | 0.622 | |
| Angry | | 0.750 | | 0.777 | |
| Nervous | | 0.962 | | 0.956 | |

However, in the three-factor model no item significantly loaded onto the third factor. Therefore, the two-factor model was adopted for further analyses. The first factor was named 'positive

emotions', and the second factor 'negative emotions'. See Table 41 for factor loadings for the six-emotion EFA models.

Mplus8 does not allow for specification of multilevel structures in EFA, so following the model building strategy suggested by Muthén (1994) the factor structure obtained for level 1 in EFA was examined for multilevel fit using multilevel CFA. Neglecting the multilevel structure of data can lead to errors in estimation (Muthén, 1991), hence this was an important step of the model development. Three-level CFA was carried out to verify the factor structure for repeated questionnaires nested within day, and days within participants. Model fit was not perfect, but still acceptable; the RMSEA and CFI indices both achieved good fit. The greatest deviation from the EFA model fit stemmed from the participant level, as evidenced by the SRMR indices (Table 42). Further analyses were conducted at the within days and between days levels, for which the model had good fit.

*Table 42. Goodness of fit indices for 3-level CFA model.*

| | | |
|---|---|---|
| RMSEA | | 0.035 |
| CFI | | 0.950 |
| SRMR | Within days (level 1) | 0.055 |
| | Between days (level 2) | 0.060 |
| | Between participants (level 3) | 0.111 |

Factor loadings for both factors at all three levels are displayed in Table 43. All factor loadings were significant at all levels. At level 2, the parameter estimate for 'calm' was greater than 1 due to a negative residual. This is not an unusual occurrence (Muthén, 1994). Constraining the residual to be positive had no effect, equating the parameter for 'calm' across levels did not change the parameter estimate either, nor did limiting the analysis to only levels 1 and 2 of the model. It is clear that there is a discrepancy between the long format data (3 levels) and the model specified. Further analyses were conducted on the wide dataset (2 levels, days nested in participants) with aggregated measures at the within-day level.

*Table 43. Standardised factor loadings for multilevel CFA.*

| | | | Estimate | S.E. | p |
|---|---|---|---|---|---|
| Level 1 | Positive emotions | Happy | 0.656 | 0.039 | *** |
| | | Relaxed | 0.590 | 0.046 | *** |
| | | Calm | 0.861 | 0.026 | *** |
| | Negative emotions | Sad | 0.478 | 0.060 | *** |
| | | Angry | 0.772 | 0.039 | *** |
| | | Nervous | 0.780 | 0.036 | *** |
| Level 2 | Positive emotions | Happy | 0.658 | 0.065 | *** |
| | | Relaxed | 0.718 | 0.070 | *** |
| | | Calm | 1.080 | 0.049 | *** |
| | Negative emotions | Sad | 0.559 | 0.070 | *** |
| | | Angry | 0.829 | 0.047 | *** |
| | | Nervous | 0.900 | 0.055 | *** |
| Level 3 | Positive emotions | Happy | 0.766 | 0.069 | *** |
| | | Relaxed | 0.563 | 0.099 | *** |
| | | Calm | 1.000 | 0.000 | *** |
| | Negative emotions | Sad | 0.576 | 0.093 | *** |
| | | Angry | 0.908 | 0.065 | *** |
| | | Nervous | 1.000 | 0.000 | *** |

Significance values *=p<0.05, **=p<0.01, ***=p<0.001

### Questionnaire-Level Analyses

The wide dataset, a two-level dataset of days nested in participants, was used to test for invariance. Due to the number of parameters in the model relative to the number of participants in the dataset, invariance testing was done separately for positive emotions and negative emotions. Factors for positive or negative emotions were specified for each questionnaire within the day (one to seven), and the three relevant emotion items loaded onto the factor. The loadings were equated across the seven factors for each emotion item; Figure 36 shows an example for the loadings of calm across questionnaires (c1-c7), onto the positive emotion factors (positive 1-positive 7), which were all equated to ‘a’.

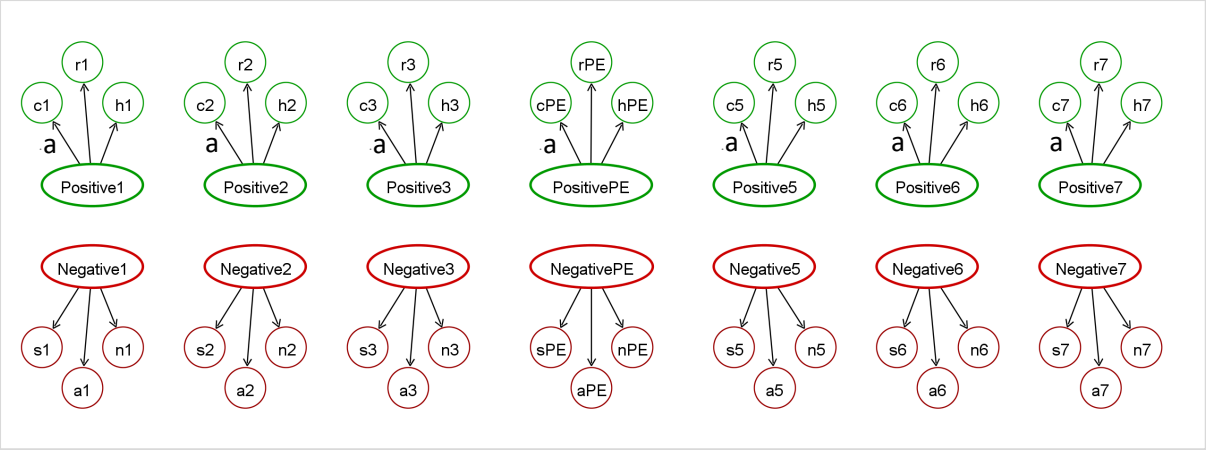



*Figure 36. Invariance testing.*

Model fit was good for positive emotions, but less so for negative emotions with only the SRMR index below the cut-off for good fit (Table 44).

*Table 44. Goodness of fit indices for invariance models.*

| | | Positive emotions | Negative emotions |
|---|---|---|---|
| RMSEA | | 0.064 | 0.087 |
| CFI | | 0.951 | 0.912 |
| SRMR | Within | 0.062 | 0.074 |
| | Between | 0.000 | 0.000 |

Using the results from the CFA, aggregated variables for positive and negative emotions were defined for each questionnaire by taking the average value of the three emotion ratings that loaded onto each factor. This was done to reduce the number of parameters for future models. Correlations were specified between timepoints within the same day, and negative and positive emotions at the same timepoint, achieving good model fit (RMSEA=0.078, CFI=0.968, SRMR within=0.100, between=0.000). Correlations were significant at $p<0.001$ between all timepoints, but both positive and negative emotions were more strongly correlated within lessons than between lessons (see Table 45 and Table 46).

*Table 45. Positive emotion correlations across the day.*

| | Positive 1 | Positive 2 | Positive 3 | Positive PE | Positive 5 | Positive 6 |
|---|---|---|---|---|---|---|
| Positive 2 | 0.83 | | | | | |
| Positive 3 | 0.72 | 0.81 | | | | |
| Positive PE | 0.39 | 0.47 | 0.49 | | | |
| Positive 5 | 0.38 | 0.41 | 0.39 | 0.58 | | |
| Positive 6 | 0.36 | 0.40 | 0.39 | 0.52 | 0.74 | |
| Positive 7 | 0.40 | 0.44 | 0.39 | 0.48 | 0.67 | 0.73 |

*Table 46. Negative emotion correlations across the day.*

| | Negative 1 | Negative 2 | Negative 3 | Negative PE | Negative 5 | Negative 6 |
|---|---|---|---|---|---|---|
| Negative 2 | 0.82 | | | | | |
| Negative 3 | 0.80 | 0.81 | | | | |
| Negative PE | 0.58 | 0.59 | 0.58 | | | |
| Negative 5 | 0.56 | 0.54 | 0.55 | 0.73 | | |
| Negative 6 | 0.61 | 0.53 | 0.51 | 0.70 | 0.78 | |
| Negative 7 | 0.59 | 0.53 | 0.49 | 0.75 | 0.80 | 0.81 |

Correlations between successive reports from the same lesson are in thick borders in tables 45 and 46: questionnaires one, two, and three were completed before PE, questionnaires five, six, and seven after PE. These form a “simplex structure” i.e., associations between reports which are located closer in time are more strongly than reports that are located further apart in time. Finally, emotions reported at the end of PE lessons were more strongly correlated with those reported during post-PE lessons than pre-PE lessons (shaded cells). There is a greater degree of stability of negative affect across all time-point, while positive affect changes more from the pre-PE classroom lesson to PE, and to the following classroom lesson.

An autoregressive model was specified, in which emotion scores at one timepoint were used to predict emotion scores at the next timepoint. As correlations between emotions in classroom lessons and PE lesson were weaker than within classroom lessons, the first model involved classroom lessons only. The second model had emotions reported at the end of PE lessons added, with autoregressions from questionnaire 3 to PE and PE to questionnaire 5, as well as direct paths from questionnaire 3 to

questionnaire 5. The direct path from questionnaire 3 to questionnaire 5 was included to test if there might be a stronger prediction of emotion states from classroom lesson to classroom lesson, compared to via the PE lesson.

Model fit of both models was close to achieving good fit (Table 47); model 2 achieved better fit on the RMSEA, whereas model 1 had better fit on the CFI, $\chi^2$, AIC, and BIC indices.

*Table 47. Goodness of fit indices for autoregressive models.*

| | | Model 1 (classroom only) | Model 2 (PE included) |
|---|---|---|---|
| $\chi^2$ (p) | | 362.74 (***) | 453.70 (***) |
| AIC | | 11981.21 | 14256.50 |
| BIC | | 12151.36 | 14309.87 |
| RMSEA | | 0.110 | 0.102 |
| CFI | | 0.912 | 0.908 |
| SRMR | Within ID | 0.100 | 0.100 |
| | Between ID | 0.000 | 0.000 |

The results of model 2 are displayed in Figure 37 (note: standardised parameters; dotted lines: p<0.05; dashed lines: p<0.01; solid lines: p<0.001). Emotions at one timepoint significantly predict emotions at the next timepoint during afternoons in school (single-headed blue arrows). For clarity of

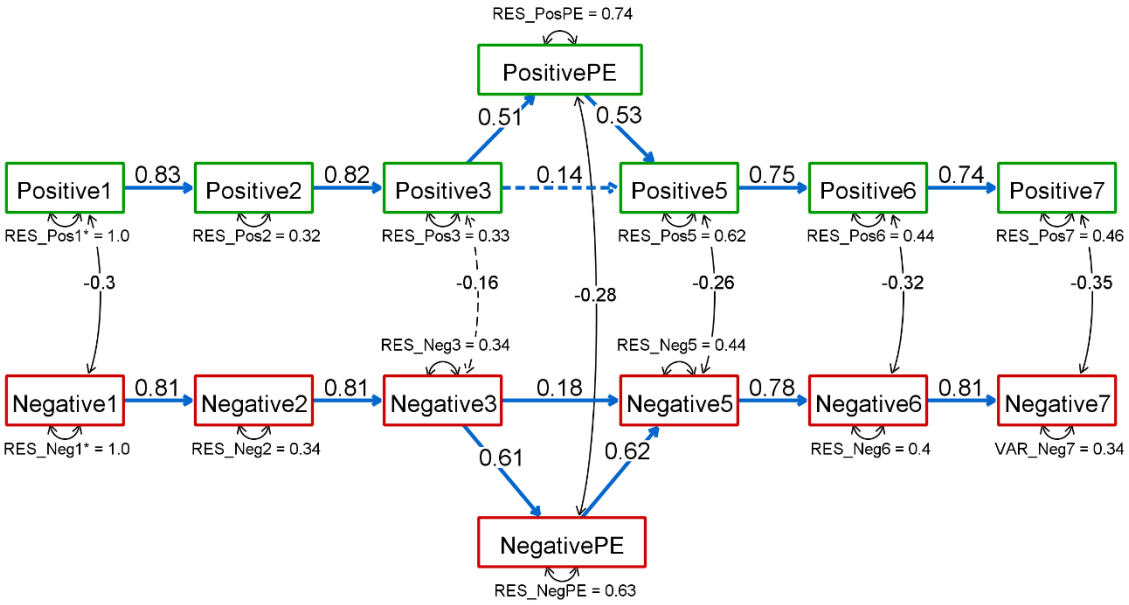



*Figure 37. Autoregression model including PE lesson (standardised results).*

the visual, non-significant paths are not shown. The direct relationship between emotions in classroom lessons before and after PE was weaker (0.14 for positive, 0.18 for negative) than the relationship between emotions in the classroom and during PE lessons (0.51 and 0.53 for positive, 0.61 and 0.62 for negative). Mediation paths from Positive3 to Positive5 via PositivePE and from Negative3 to Negative5 via NegativePE were added to the model and were significant ($\beta_{positive}$=0.28, SE=0.03, p<0.001, $\beta_{negative}$=0.44, SE=0.04, p<0.001), meaning that indirect paths were stronger than direct paths.

### Lesson-Level Emotion Models

To enable further models of emotions scores with lesson-level predictors, and as correlations and autoregressions indicated strong relationships between within-lesson questionnaires, a correlation model was specified for emotions at the lesson level. This model is depicted in Figure 38 (note: standardised parameters; dotted lines: p<0.05; dashed lines: p<0.01; solid lines: p<0.001). This model had good model fit (RMSEA: 0.060, CFI:0.959, SRMR:0.047). Positive post-PE emotions were more strongly correlated with PE lesson positive emotions than pre-PE positive emotions. Post-PE negative emotions on the other hand were more strongly correlated with pre-PE negative emotions than PE lesson negative emotions.

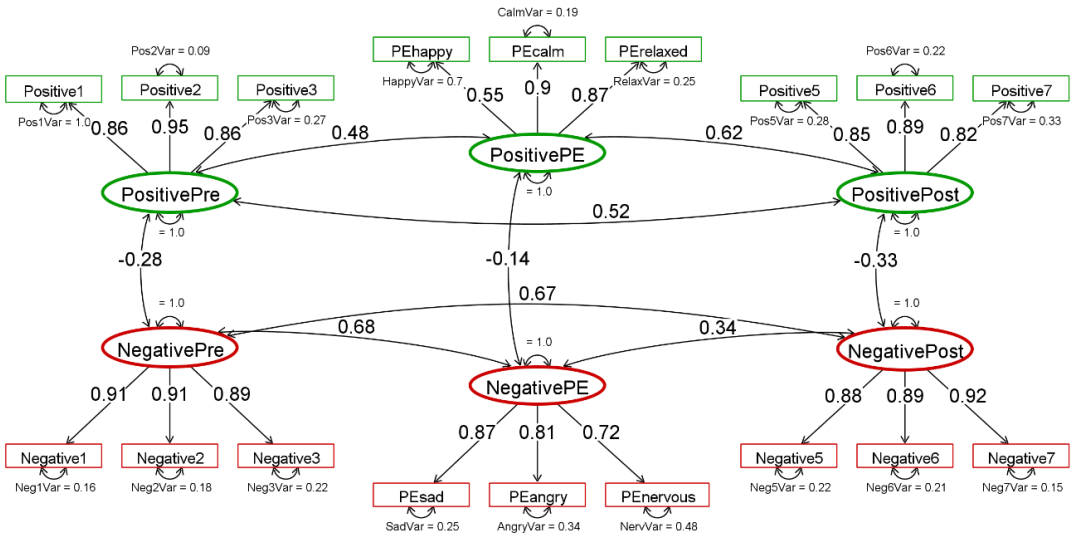



*Figure 38. Correlation at lesson level (standardised results).*

This lesson-level structure was applied to a regression model to investigate the predictive effects of pre-PE and PE lesson emotions on post-PE emotions, displayed in Figure 39 (note: standardised parameters; dotted lines: p<0.05; dashed lines: p<0.01; solid lines: p<0.001).

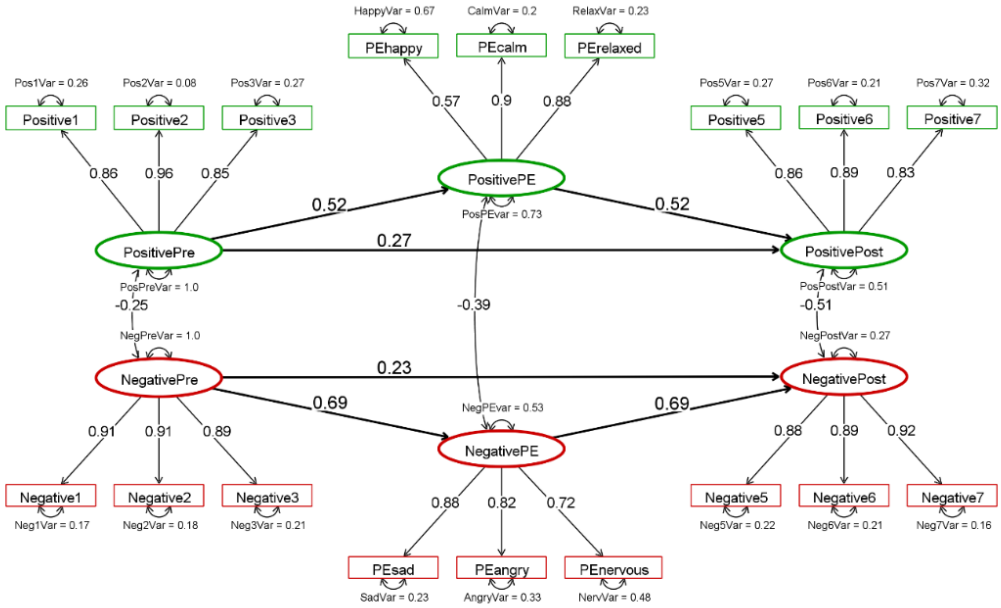



*Figure 39. Regression of emotion state at lesson level (standardised results).*

All regression paths and correlations were significant at p<0.001, and model fit was good (RMSEA:0.065, CFI:0.947, SRMR (between):0.078). Regression coefficients for both positive and negative emotions were stronger from pre-PE to PE and PE to post-PE than directly from pre-PE to post-PE. Mediation paths were added to the model and found to be significant for both positive and negative emotions ($\beta_{positive}$=0.30, SE=0.04, p<0.001, $\beta_{negative}$=0.55, SE=0.06, p<0.001). Mediation paths were also stronger than direct paths, especially for negative emotions ($\beta_{direct}$=0.23, $\beta_{indirect}$=0.55).

Cross-lagged paths between positive and negative emotions were added to allow for the possibility of positive and negative emotions to influence one another. Adding the cross-lags improved model fit on both the relative and absolute indices (Table 48). Cross-lags were found to be significant, but with small coefficients. Figure 40 displays the model with cross-lags (note: standardised parameters; dotted lines: p<0.05; dashed lines: p<0.01; solid lines: p<0.001).

*Table 48. Goodness of fit indices for autoregression model with and without cross-lags.*

| | | Model 1 (without cross-lags) | Model 2 (with cross-lags) |
|---|---|---|---|
| $\chi^2$ (p) | | 410.88 (***) | 359.79 (***) |
| AIC | | 19864.10 | 19821.01 |
| BIC | | 20133.89 | 20107.92 |
| RMSEA | | 0.065 | 0.060 |
| CFI | | 0.947 | 0.955 |
| SRMR | Within ID | 0.078 | 0.051 |
| | Between ID | 0.000 | 0.000 |

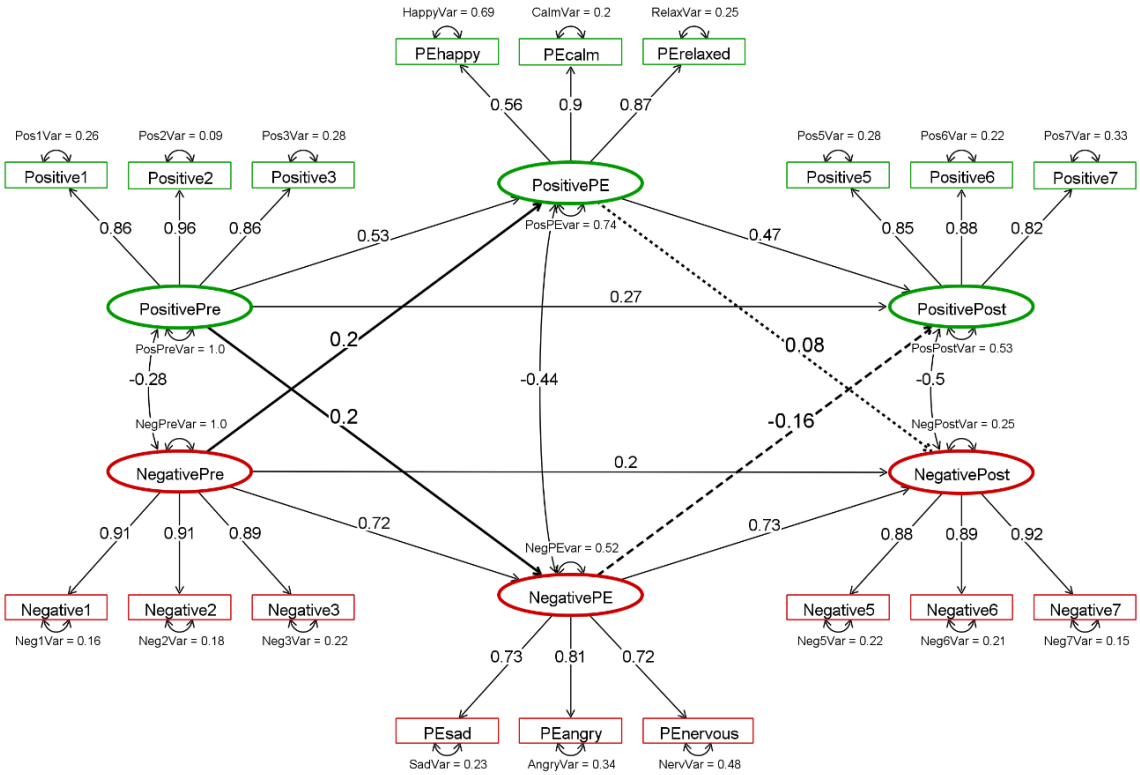



*Figure 40. Regression of emotions at lesson level, with cross-lags (standardised results).*

In Figure 41, the mediation paths from pre-PE positive to post-PE positive via PE positive are highlighted in green (top half bold paths), and from pre-PE negative to post-PE negative via PE negative in red (bottom half bold paths). Both remained significant ($\beta_{positive}$=0.27, SE=0.04, p<0.001, $\beta_{negative}$=0.63, SE=0.07, p<0.001) and exceeded the direct paths (note: standardised parameters; dotted lines: p<0.05; dashed lines: p<0.01; solid lines: p<0.001).

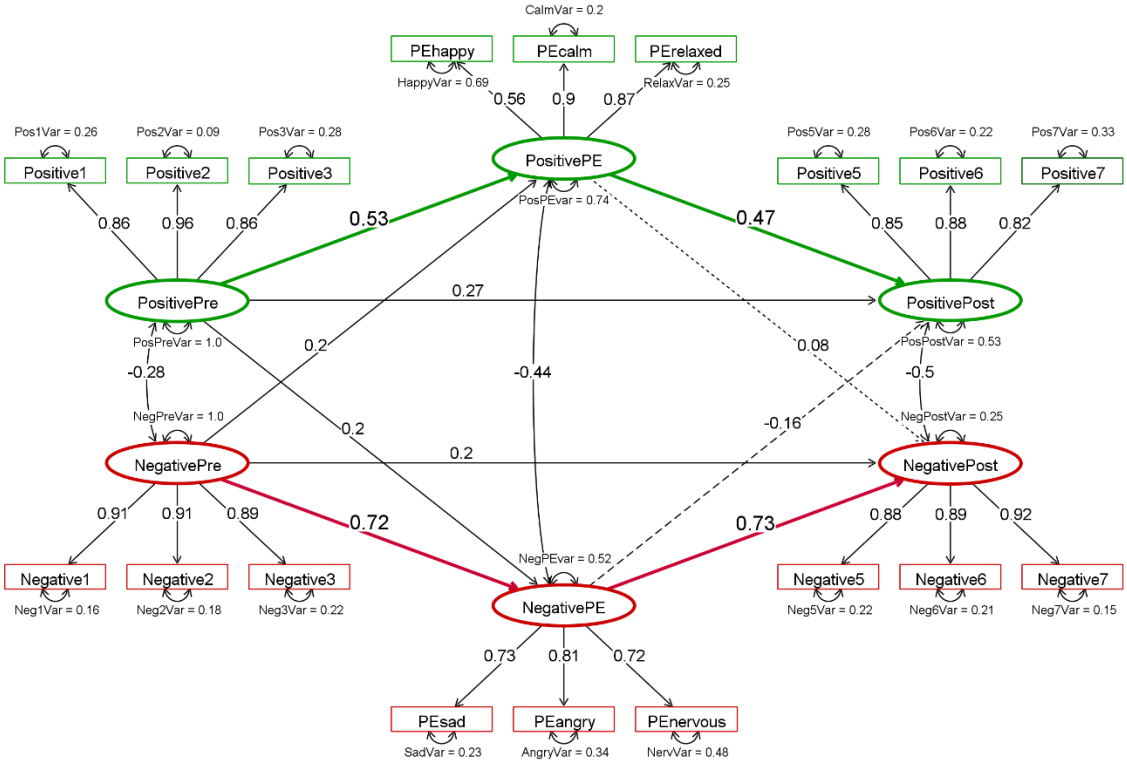



*Figure 41. Regression with cross-lags and mediation paths (standardised results).*

### Emotion Models with Predictors

PE lesson content was added to the model as a predictor of emotions at the end of PE lessons. Model fit changed slightly, with goodness of fit according to the $\chi^2$, AIC, and BIC declining slightly and the RMSEA, CFI, and SRMR improving slightly (Table 49).

*Table 49. Goodness of fit indices for emotion state model with and without predictors.*

| | | Model 1 (without predictors) | Model 2 (with PE condition) | Model 3 (with PE condition and PE enjoyment) |
|---|---|---|---|---|
| $\chi^2$ (p) | | 359.79 (***) | 433.17 (***) | 542.84 (***) |
| AIC | | 19821.01 | 19824.50 | 21529.66 |
| BIC | | 20107.92 | 20154.24 | 21898.90 |
| RMSEA | | 0.060 | 0.046 | 0.052 |
| CFI | | 0.955 | 0.957 | 0.941 |
| SRMR | Within ID | 0.051 | 0.046 | 0.052 |
| | Between ID | 0.000 | 0.000 | 0.000 |

Two of the intervention conditions were found to have a significant effect on emotion state at the end of PE lessons: medium intensity/low complexity (β=0.12, SE=0.05, p<0.05) and medium intensity/high complexity (β=0.16, SE=0.05, p<0.01), but on negative emotions only. Figure 42 shows the model with PE lesson content as predictors (note: standardised parameters; dotted lines: p<0.05; dashed lines: p<0.01; solid lines: p<0.001). Mediation paths from PE lesson content to post-PE emotions revealed significant effects of both medium-low and medium-high PE lessons on post-PE negative emotions (blue paths, $\beta_{med\text{-}low}$=0.24, SE=0.10, p<0.05, $\beta_{med\text{-}high}$=0.31, SE=0.10, p<0.01).

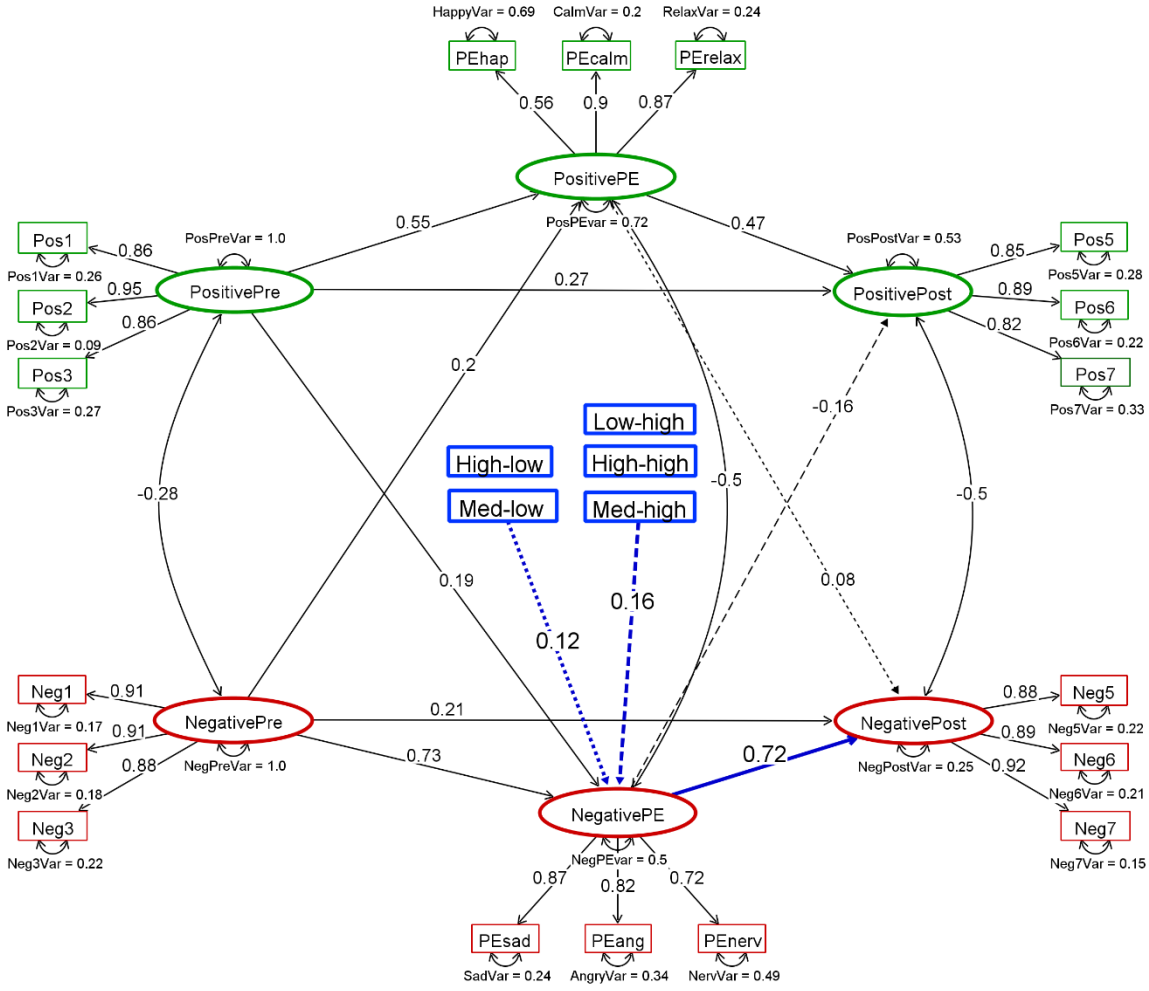



*Figure 42. PE lesson content predicting PE emotions (standardised results).*

Next, PE lesson enjoyment was added as a predictor of emotion states at the end of PE. Model fit worsened (Table 49), despite the finding that PE enjoyment was significantly predictive of both positive and negative emotions ($\beta_{positive}$=0.31, SE=0.04, p<0.001, $\beta_{negative}$= -0.20, SE=0.04, p<0.001). PE

enjoyment also had a significant indirect effect on both positive and negative emotions in the classroom after PE lessons ($\beta_{positive}$= 0.15, SE=0.03, p<0.001, $\beta_{negative}$= -0.15, SE=0.03, p<0.001). These results are displayed in Figure 43, with blue paths showing the mediation of PE enjoyment by negative emotions, and orange paths the mediation by positive emotions (note: standardised parameters; dotted lines: p<0.05; dashed lines: p<0.01; solid lines: p<0.001). Yet, PE lesson content did not significantly predict PE lesson enjoyment.

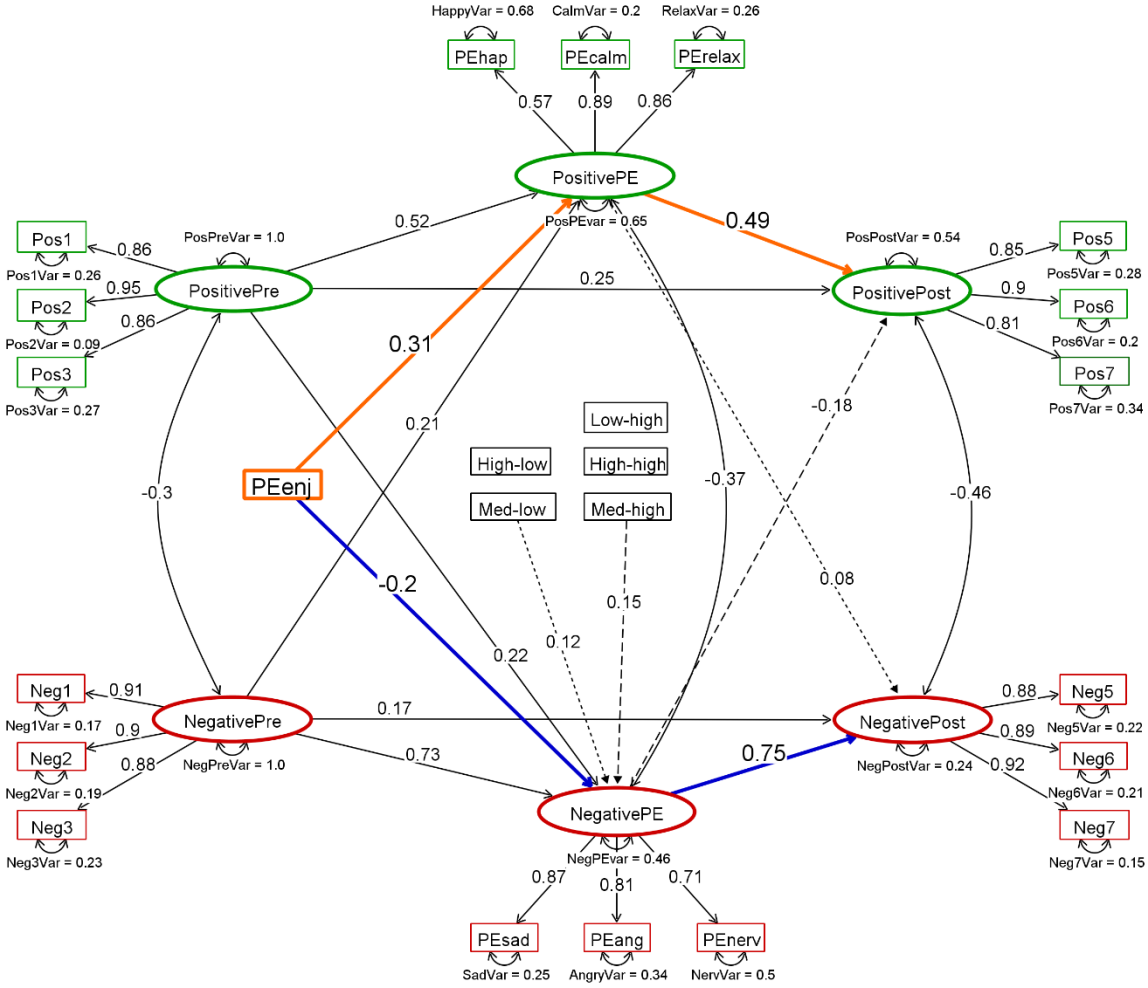



*Figure 43. PE lesson content and enjoyment predicting emotions (standardised results).*

### RQ5: Mediation of Physical Activity and On-Task Behaviour via Emotion States

On-task behaviour in the classroom was added to the model as the outcome variable, changing post-PE emotion states into mediation variables. Model fit improved relative to the models with emotion state as the outcome, although absolute fit improved by only a small amount (Table 50).

*Table 50. Goodness of fit indices for model with and without behaviour outcomes.*

| | | Model 1 (without behaviour outcomes) | Model 2 (with behaviour outcomes) |
|---|---|---|---|
| $\chi^2$ (p) | | 542.84 (***) | 598.05 (***) |
| AIC | | 21529.66 | 18558.95 |
| BIC | | 21898.90 | 18660.61 |
| RMSEA | | 0.052 | 0.051 |
| CFI | | 0.941 | 0.937 |
| SRMR | Within ID | 0.052 | 0.052 |
| | Between ID | 0.000 | 0.000 |

PE lesson content directly predicted post-PE on-task behaviour for medium intensity/low complexity ($\beta_{medium\text{-}low}$=0.14, SE=0.06, p<0.05), low intensity/high complexity ($\beta_{low\text{-}high}$=0.14, SE=0.06,

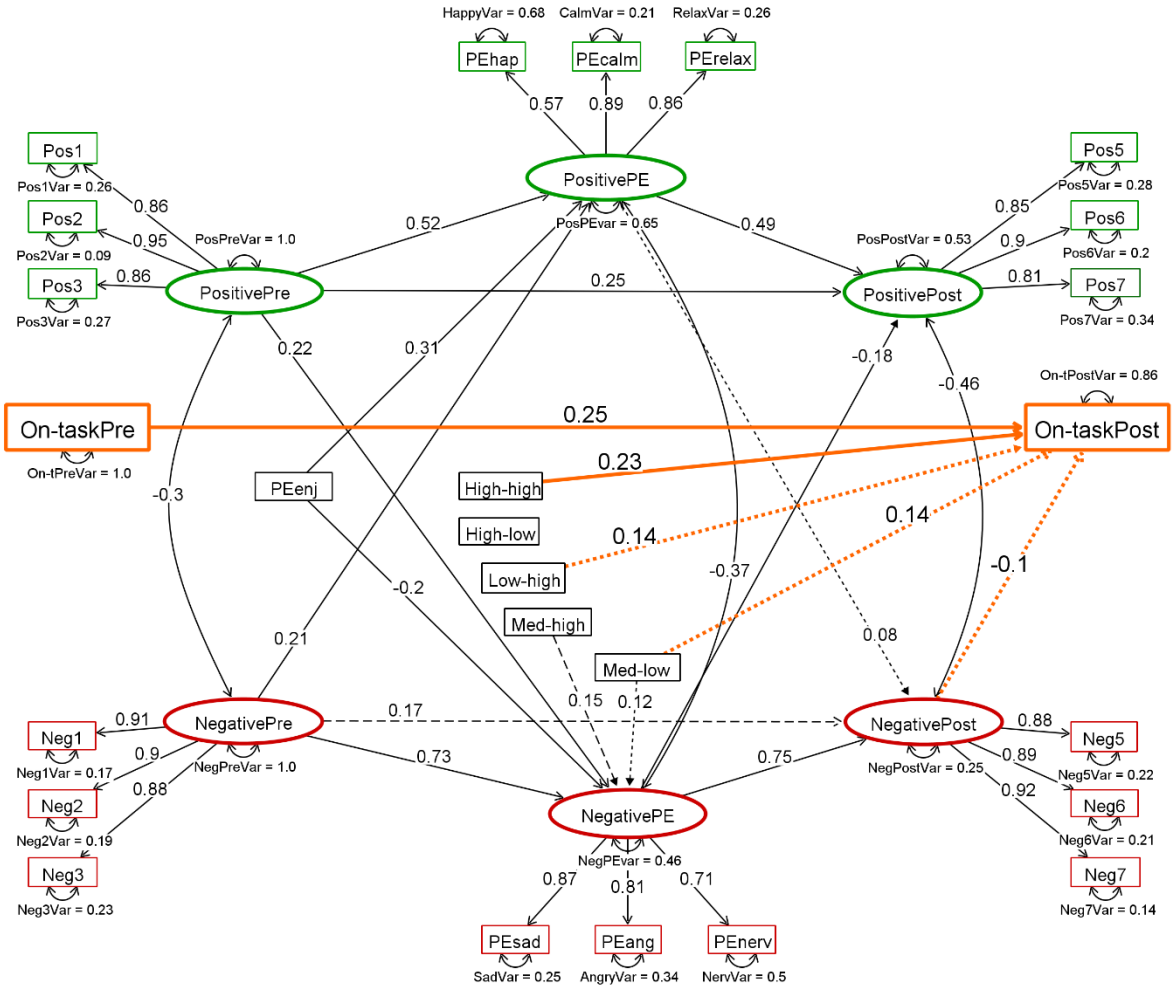



*Figure 44. PE lesson content, emotions, and task-behaviour (standardised results).*

$p<0.05$), and high intensity/high complexity ($\beta_{high\text{-}high}=0.23$, SE=0.05, $p<0.001$) lessons.

The final mediation model is displayed in Figure 44. Pre-PE, neither positive nor negative emotions predicted on-task behaviour. On-task behaviour after PE was predicted only by negative emotions ($\beta_{negative}=-0.10$, SE=0.05, $p<0.05$), with less on-task behaviour as a result of increased negative emotions. Mediation paths from PE conditions via PE emotions and post-PE emotions to post-PE on-task behaviour were non-significant (note: standardised parameters; dotted lines: $p<0.05$; dashed lines: $p<0.01$; solid lines: $p<0.001$). Thus, the hypothesis that there is a mediation effect of PA on on-task behaviour via emotional state was rejected.

## 4.5 Concluding Remarks

The results presented in this chapter have provided answers to the research questions. MLM showed no main effect of either PE lesson complexity (RQ1, low complexity: d=0.11, OR=1.22; high complexity: d=0.16, OR=1.34) or intensity (RQ2, low intensity: d=0.11, OR=1.22; medium intensity: d=0.11, OR=1.22; high intensity: d=0.19, OR=1.41) on subsequent on-task behaviour in the primary school classroom. An interaction of complexity and intensity was found (RQ3, high intensity/high complexity: d=0.33, OR=1.82), but the direction of the interaction effect was not as hypothesised. The hypothesised inverted-U relationship for low complexity lessons was not significant (low intensity/low complexity: d=0.09, OR=1.25; medium intensity/low complexity: d=0.18, OR=1.40; high intensity/low complexity: d=d=0.05, OR-1.10), and at high complexity not all lessons were more efficient than at low complexity (medium intensity/low complexity: d=0.18, OR=1.40, medium intensity/high complexity: d=0.04, OR=1.07). Individual differences in the response to the PE intervention were found (RQ4), but again the direction of these effects was not always as hypothesised. As expected, those below the 33$^{rd}$ percentile of on-task behaviour before PE significantly benefited from all intervention conditions (d=0.30-0.72, OR=1.70-3.70), whereas the whole sample only significantly benefited from the high intensity/high complexity condition. But, contrary to the hypothesis, those with the lowest attainment

levels benefited more than those achieving above age-related expectation (d=0.17-0.52 vs. d= -0.24-0.22, OR=1.35-2.55 vs. OR=0.65-1.49). Gender differences were also opposite to those hypothesised, as girls benefited from three conditions (d=0.20-0.43, OR=1.43-2.16), but boys didn't significantly benefit from any. Finally, no mediation effect of emotion states on the effect of PA intensity and complexity on subsequent on-task behaviour was observed (β= -0.011-0.002, p=0.070-1.000).

In the next chapter, the results will be discussed in the context of, and related and compared to, the current body of literature. The behaviour results will be positioned in the landscape of recent research conducted in the field of PA and cognition, with a particular focus on those studies conducted in schools. Mediation models will be compared to work taken from the emotional and behavioural engagement literature.

# 5 Discussion

In this chapter, the results from the analyses presented in chapter 4, and the answers to the research questions will be discussed in relation to the literature. Next, implications for practice and policy are highlighted. Finally, strengths and limitations are identified and suggestions made for future research.

## 5.1 Research Question 1

It was investigated if exercise complexity has a main effect on subsequent on-task behaviour. It was hypothesised that high complexity PA would have a greater beneficial effect than low complexity PA. However, no main effect was found, and the hypothesis for the first research question was rejected.

De Greeff *et al.* (2018) concluded in their meta-analysis that although there was a small-to-moderate effect of acute PA on selective attention (Hedges' $g=0.43$, $p<0.05$), there was no significant effect of cognitively engaging PA. However, only five studies were included in the meta-analysis and results from studies in this field are still inconsistent (Mavilidi *et al.*, 2018), possibly due to variation in a range of parameters. Hence it is important to consider specific aspects of these studies more closely.

The effects of cognitively engaging PA on attention seem to vary by participant age. Studies involving children aged up to eleven have mostly reported no effects of cognitively engaging PA over simple PA. In a lab-based study of six- to ten-year old children taking part in 20 minutes of exergaming, Best (2012) reported that participants' performance on the flanker task showed no improvement in measures of attention as a result of added cognitive engagement in the exergaming task. Likewise, Egger, Conzelmann and Schmidt (2018) found no effect of added cognitive engagement on inhibition in a 20-minute intervention in a sample with a mean age of 7.94 years, and Gallotta *et al.* (2015) reported negative effects of cognitive engagement during PE lessons on subsequent measures of

attention in the d2 task in children aged eight to eleven. On the other hand, studies in older samples (adolescents aged eleven to sixteen) have reported positive effects (e.g. Schmidt *et al.*, 2016; Budde *et al.*, 2008). This was also concluded by Chang *et al.* (2012); in their meta-analysis they found greater effect sizes for older adolescents, adults and elderly samples compared to younger children. This study involved participants aged seven to ten, which may have contributed to the effect sizes being as small as they are.

There is a wide variety in the methods by which researchers operationalise cognitive engagement in PA interventions. In this study, cognitive demands were included in the PA by increasing the demands from coordination, reaction time, movement sequencing and timing. Studies using this method of increasing cognitive demand have reported positive effects on measures of attention, such as Budde *et al.* (2008), who found that d2 task performance in thirteen- to sixteen-year olds improved after ten minutes of coordinative exercise compared to simple exercise. On this basis, a main effect for complexity was expected. It is possible though, that the use of a d2 task immediately after PA, rather than observations of behaviour over a 25-minute period, has contributed to the disparity in results.

The results for complexity in this study are similar to studies which have applied a different approach to adding cognitive engagement to PA: by repeatedly manipulating the rules of the activity, or by increasing the number of simple elements or movements used to make up an activity. Studies using these approaches have reported no effects of complex PA on measures of attention. For example, after participation in exercise involving simple movements with added complexity from changing the instruction mapped to each movement and increasing the number of movements used, Egger *et al.* (2018) found no improvement in inhibition scores in a flanker task. In a comparable design, but over a 20-week period, Egger *et al.* (2019) identified no effect of cognitive engagement in PA on measures of inhibition in the flanker task. Similarly, Best (2012) reported no improvement in measures of attention on the d2 task after playing exergames with cognitive engagement induced by increases in the range of simple movements needed to complete the game, as well as the introduction of

reaction time and decision-making by requiring participants to adjust their motor responses based on random and unpredictable visual stimuli in the game. In this study, the main reason no effect of complexity was found, was the lack of an effect in the medium intensity-high complexity condition. This PE lesson consisted of paired and small-team games, relying on reacting to the actions of others. Although another source of complexity in this lesson was the use of the non-dominant hand for throwing and catching, it is possible that the need to react to visual or verbal cues to perform motor skills has a different effect on measures of cognition than the complexity inherent in performing coordinative exercise.

In another approach which has yielded positive results for cognitively challenging PA, curriculum content is overlaid on simple movements, often in programmes taking place in classroom environments. For example, positive effects on on-task behaviour were reported by Mavilidi *et al.* (2019) after participation in a five-minute intervention for both the simple and cognitively engaging PA groups, but with greater improvements for the group taking part in PA overlaid with maths lesson content. Also using a maths game in the cognitively engaging ten-minute PA intervention, Schmidt *et al.* (2016) reported main effects of cognitive engagement on focussed attention in the d2 test. In this approach, the cognitive challenge and PA performed are not necessarily linked, and it is imaginable that this type of cognitive stimulation more readily transfers to classroom-based measures of cognition or the strategies required for completing cognitive tasks in the classroom.

Another point to note is that the majority of positive results for cognitive engagement in PA have been reported for interventions with a short duration of up to ten minutes. Mavilidi *et al.* (2019), Schmidt *et al.* (2016), and Budde *et al.* (2008) all reported on interventions of between five and ten minutes, and all found positive effects for cognitive engagement. Several longer interventions have either found no effects, or even negative effects of cognitive engagement. After a 20-minute intervention, (Egger *et al.*, 2018) found no effects of cognitive engagement on measures of attention, and after 50-minute interventions involving full PE lessons, Gallotta *et al.* (2015) identified negative effects of cognitive engagement during PE on performance in the d2 test of attention.

Gallotta *et al.* (2015) suggest that the long duration of the intervention might have over-stretched pupils' cognitive capacity and therefore negatively affected subsequent attentional performance. The literature reports on studies which present PA with or without cognitive engagement, but unlike PA intensity which is divided into sedentary, low, moderate, and vigorous, the only distinction made for cognitive engagement is high or low. Similar to the effects of PA being influenced by the level of intensity and type of PA, it is feasible that the effects of cognitive engagement also depend on the level and type of cognitive engagement. The notion of 'too much cognitive engagement' leading to a cognitive overload was also proposed by Egger *et al.* (2018), although they related this to individual variations in cognitive capacity and the need to adjust intervention content to it. Considering the previous points highlighted from the current literature regarding the age of participants (younger participants tending to respond less favourably) and the type of cognitive engagement (most favourable being curriculum content mapped onto PA, or isolated coordinative exercises), it is not unreasonable to think that these previously mentioned differences in effects could be related to the individual's cognitive capacity; older children having greater cognitive capacity, and lower demands involved in simple exercise combined with curriculum content or isolated coordinative exercises.

The participants' motor skill level could be a mediator in the effects of exercise complexity. Based on a participant's skill level, the cognitive demand induced by the complexity of PA in the intervention lessons will have varied (Singh *et al.*, 2019). As time and budget did not allow for the testing of motor skills in this study's sample, it could not be included in the analyses as a control or mediation variable, and effects of complexity may therefore have gone undetected.

It is reasonable that cognitive overload could have happened in the study discussed here, due to the duration of the intervention lesson. And, particularly in the medium intensity-high complexity lesson the cumulative effect of complexity from coordinative exercise for successful object manipulation with the non-dominant hand, object manipulation whilst moving, strategic planning of movements in the competitive team games setting, and the cooperative work required within the teams might have been too much to cope with. In the low intensity-high complexity setting the

cognitive demands were based on complex object manipulation, with some exercises involving reaction time. In the high intensity-high complexity lesson, cognitive demand was created by the simultaneous coordination of multiple limbs, and the need to be in time with music. This was however modelled by the lead researcher throughout, so for participants struggling to remember the sequence or to execute the movements successfully, there was some support available to keep them engaged.

Finally, the lack of an effect for PA complexity could be a result of the difference in measures used to assess the effects. In the majority of published studies, attention was assessed with the d2 task or flanker task, completed immediately after exercise in a laboratory setting or under exam-like conditions in a school environment, whereas in this study the behavioural manifestation of the hypothesised increase in attentional capacity was examined. It could therefore be the case that any potential increases in attentional capacity were not detected above and beyond the melee that forms the daily goings-on in a primary school classroom.

In conclusion, the lack of an effect identified for PA complexity could be the result of a number of factors. Participant age, the type and combination of cognitive demands involved in the lessons, the motor skill level of the participants, and the duration of the lessons leading to overload have all been discussed in this section. More systematic research is needed to investigate if there is a dose-response relationship between the level of cognitive demand present in the PA and subsequent measures of attention. Moreover, additional research combining measures of attention derived from cognitive tests and behavioural manifestations of these measures would advance the knowledge in this field.

## 5.2 Research Question 2

Next, it was examined if there was a main effect of exercise intensity on subsequent on-task behaviour. It was hypothesised that exercise intensity would have an inverted-U relationship with subsequent on-task behaviour, thus the greatest benefit would be for medium intensity PA. No main effect was identified. Contrary to expectation, high intensity exercise was the only one getting close

to a small effect size (d=0.19), whereas the effect size for both low and medium intensity exercise was d=0.11.

Especially the lack of effect in the moderate intensity condition was surprising. Chang *et al.* (2012) conducted a meta-analysis of the effects of various intensities of acute PA on attentional performance in humans of all ages. They found that immediately after PA, only LPA to MPA had an effect, whereas after a delay, all but very light intensity PA had positive effects. On this basis, both the low and medium intensity conditions should have recorded effects, and VPA should not. Hence, no effect of VPA was predicted, and to find the greatest effect size for the high intensity conditions was unexpected. Yet, similar to the pattern of effects pertaining to high intensity PA in this study, Grieco *et al.* (2016) identified greater benefits of active breaks compared to sedentary breaks, and far greater benefits of MVPA breaks than LMPA breaks. However, they do caution that the group taking part in MVPA had substantially lower baseline time-on-task, and the effect for MVPA may have been inflated due to this. In this study, the level of on-task behaviour before PE did differ between the intensity levels, but high intensity had the greatest effect size, despite having the highest level of pre-PE on-task behaviour.

Contrary to the findings of this study, but in line with findings from Chang *et al.* (2012) for effects immediately after PA, Pellegrini and Davis (1993) found that greater duration of VPA during break-time corresponded with reduced time-on-task after break. However, Lambourne and Tomporowski (2010) concluded in their meta-analysis that the debilitating effects of VPA on cognition have not been conclusively evidenced in laboratory settings. They did find convincing evidence of the positive effect on moderate intensity exercise on cognition, which was elusive in this study, although no specific mention was made of attentional allocation (ibid.).

Despite the overall positive effects reported in several meta-analyses and systematic reviews (e.g. Chang *et al.*, 2012; Lambourne and Tomporowski, 2010), including some specifically of the literature on children's response to PA (e.g. Donnelly *et al.*, 2016; Janssen *et al.*, 2014), a range of studies with samples including children between the ages of six and twelve have reported no effects of PA on measures of attention. Although many of these studies compared intervention conditions of PA with

sedentary control groups, rather than the effects of varying levels of PA, this is still relevant to the research question here. In this study no effect of PA overall was identified (the difference from pre- to post-PE) above the d=0.2 cut-point for a small effect size, and this research question merely highlighted that this absence of an effect did not differ by intensity level of the PA.

A frequently used test for measures of attention in studies of PA is an adapted flanker test for use with child samples (e.g. Egger *et al.*, 2019; Best, 2012). This is an established test, and there are studies which have used this test alongside other measures of EF, such as shifting (e.g. Schmidt, Egger and Conzelmann, 2015a) or updating (e.g. Jäger *et al.*, 2015) where those measures did improve, but attention did not, even when tested in the same study with the same sample of children. It is therefore possible that measures of attention are less sensitive to the influences of PA than some other cognitive measures.

It is also possible that the effects of cardiovascular exercise depend more greatly on chronic participation over time. Relationships between fitness and improvements in functional brain structure in the prefrontal cortex and cerebellum have been indicated as a possible link in the relationship between long-term participation in cardiovascular training and cognitive function (Voelcker-Rehage and Niemann, 2013). It is thought that changes in brain function as a result of long-term high levels of PA may be because greater blood vessel density leads to better oxygen supply to neurons (ibid.). In light of this, aerobic fitness would be an important variable to include in analyses. Unfortunately, the budget and timeline for this study did not allow for measures of aerobic fitness to be collected.

In conclusion, the lack of a main effect for PA intensity may be due to measures of attention not being equally sensitive to the influence of PA overall as some other EF. Alternatively, as effect sizes are below the d=0.2 cut-point across all intensity levels, it may be that, similar to PA complexity effects, the effects of intensity were simply too small to be detected above and beyond all the contextual influences in the classroom. Another possible reason is that aerobic fitness was not controlled for in the analyses. Lastly, the lack of an effect may be masked by the interference of the high and low levels of complexity at each intensity level, which may be interacting with the intensity level in such a way

that opposing effects at high and low complexity are annulling each other. This is further explored in section 5.3, which discusses the interaction effects of PA intensity and complexity.

## 5.3 Research Question 3

An interaction between complexity and intensity was hypothesised. It was expected that in low complexity lessons, intensity would have an inverted-U relationship with subsequent on-task behaviour, and high complexity would have a linear relationship. High complexity was predicted to have a greater effect than low complexity at each intensity level. An interaction effect of complexity and intensity was confirmed, but the nature of the hypothesised dose-response was not. At low complexity, an inverted-U relationship was found, but effect sizes for all three conditions were below d=0.2. Medium intensity did approach the cut-off (d=0.18), whereas effect sizes for both low and high intensity did not reach d=0.1. High complexity lessons had greater effects for low and high intensity (d's low intensity: 0.09 and 0.12, d's high intensity: 0.05 and 0.33), but at medium intensity the effect of high complexity was smaller than at low complexity (d's medium intensity: 0.18 and 0.04). Only for high intensity-high complexity did the effect size exceed the cut-off for a small effect (d=0.33).

To start with, the results of the low complexity conditions will be discussed. The hypothesised inverted-U relationship was identified for low complexity exercise, but effect sizes for all intensities failed to reach d=0.2, thereby lending no practical significance to these findings. This may be explained by the contextual factors involved in real-life settings in the classroom, as discussed previously in sections 2.3 and 3.3.4. This was suggested also by Jäger *et al.* (2015), who argued that effects may be harder to attain in real-world settings than in laboratories.

Next, the identified relationship between complex PA and intensity will be discussed. This was different from the hypothesised relationship insofar that unexpectedly medium intensity lessons at high complexity had a smaller effect than at low complexity. This meant that the predicted linear relationship was not confirmed. At low and high intensity, the interaction with complexity did follow

the hypothesised direction as effect sizes were greater at high complexity than at low complexity. Yet, the effect size of low intensity-high complexity lessons fell short of d=0.2 (d=0.12).

Grieco *et al.* (2016) investigated the effects of cognitively engaging breaks at various intensities on time-on-task and established a positive effect of increasing PA intensity in ten- to fifteen-minute long cognitively challenging classroom breaks. Sedentary breaks had no effect (d=0.06), LMPA achieved an effect size of d=0.43, and MVPA of d=1.22. In contrast to Grieco *et al.* (2016), the findings of the study discussed here indicated increased on-task behaviour only after the high intensity-high complexity lesson, and not after the low or medium intensity with complexity. The difference in duration of the intervention may have contributed to this difference in findings, as children participated in around 45 minutes of PA in the current study, almost three times as long. Duration of PA has been previously identified as a possible predictor of effectiveness (Chang *et al.*, 2015), and activity breaks of ten minutes were found most effective by Howie *et al.* (2014). Another possibility is that the intensity of PA in the low intensity lesson was below the LMPA condition in the study by Grieco *et al.* (2016), and therefore failed to reach an effect size above d=0.2.

Several previous studies have reported positive interaction effects of MPA or MVPA and cognitive engagement. For instance, Jäger *et al.* (2014) found interaction effects of MVPA and cognitive engagement on measures of inhibition in a modified flanker task for children, despite finding no main effects for PA or cognitive engagement. Their intervention was delivered to groups of four children at a time, rather than to whole classes of children. Mavilidi *et al.* (2019) delivered their intervention at a class level, and also identified an interaction of MPA and cognitive engagement for five-minute active classroom breaks with mathematics content. They found that on-task behaviour in nine-year-olds increased significantly after simple PA breaks, but even more so after cognitively engaging PA. Although these results indicate that PA and cognitive engagement interact, they were achieved by comparing MVPA with and without cognitive engagement to no PA with and without cognitive engagement, rather than comparing to varying levels of PA intensity.

The lack of an effect of the medium intensity lessons with high complexity may be related to the difficulty of detecting effect sizes in real-world settings, as mentioned previously, or a lower sensitivity of measures of attention leading to inconsistent results. Not all studies investigating the interaction of PA and cognitive engagement found effects on attention. Best (2012) found no main or interaction effects of MPA and cognitive engagement in six- to ten-year-olds, and the same was the case for Jäger *et al.* (2015) in slightly older children, aged between ten and twelve. In children aged eleven and twelve, Schmidt *et al.* (2016) identified main effects of cognitive engagement on focused attention, but no interaction with MPA. Again, these studies compared children taking part in MPA to those in sedentary conditions, and this must be borne in mind when comparing their results to this study. Also, the notion of overload must once more be considered in the case of this particular intervention condition, which contained many elements of cognitive complexity. Finally, there is the possibility that sample level effects are hiding effects of PA for certain groups or individuals. Although several participant-level variables were controlled for in the multilevel analyses, further investigation of the interaction effects of PA exposure and individual characteristics did uncover different relationships than sample-level investigations. These individual differences are discussed in the following section.

## 5.4 Research Question 4

Individual participants were predicted to respond differently to PA based on their proportion of pre-PE on-task behaviour, their attainment level, their sports participation habits, and their gender. Individual differences were identified in participants' response to PA of varying intensities and complexity, but the hypothesised direction of the relationship between participant characteristics and PA response was confirmed only for pre-PE proportion of on-task behaviour.

### 5.4.1 Pre-PE On-Task Behaviour

Those who displayed the lowest proportion of on-task behaviour pre-PE were predicted to benefit to a greater extend from PA participation than those with higher proportions pre-PE on-task behaviour. This hypothesis was confirmed for all intervention conditions. In this study the comparison was drawn between the full sample and those below the 33$^{rd}$ percentile of the sample. The greatest effect size was found for the high intensity-high complexity condition (d=0.72), but this was still a way off the effects identified by Mahar *et al.* (2006) (d=2.2). Mahar *et al.* (2006) defined the 'least on-task' students as those who spent less than 50% of the pre-break lesson on-task, compared to this study using a cut-off based on the bottom third of the sample. This corresponded with pre-PE proportions on-task of less than 0.68. It was deemed that using an absolute rather than a relative cut-off, such as done by Mahar *et al.* (2006), or only those in the bottom quartile, would have led to too small a sample size in the 'low on-task' group to draw a meaningful conclusion. Bearing in mind the negative covariance between the intercept and slope identified in this study, it is not hard to imagine that by setting the cut-point for 'low on-task' even lower, the effect size for this group would increase.

Although the sub-sample categorised as the lowest on-task were only one third of the total sample size, and results from this analysis must be interpreted cautiously, this is a reflection of the real situation in schools. Those with very low levels of on-task behaviour are a minority, yet they are a minority which causes the greatest disruption to instructional time and not only their own learning, but also the learning of others in their class (Mahar *et al.*, 2006; Godwin and Fisher, 2011). It is therefore important for teachers to be equipped with the tools to tackle this behaviour, and PA appears to be one method to contribute to the solution.

### 5.4.2 Attainment

In contrast to Schmidt *et al.* (2015b), who reported no effects of attainment on their findings, the effect of five of the intervention conditions in this study differed by the participant's level of attainment. The only exception was the medium intensity-high complexity lesson, which failed to

achieve an effect size beyond d=0.2 regardless of participants' attainment. No clear pattern of dose and response was apparent in the interaction effect of PA and attainment, but unlike Jäger *et al.* (2015), who found that only those with attainment scores in the top half of the sample experienced positive effects of PA, the greatest benefits of PA in this sample were achieved by those with the lowest attainment. In fact, the highest achieving group in this study were the only participants to attain one small negative effect, and one negative effect below the d=-0.2 threshold. These were in the high intensity-low complexity lesson (sprinting races), and the medium intensity, high complexity lesson (team games) respectively. These were the only two lessons involving elements of competition against others, although no scores were recorded or made explicit by the lead researcher during the lesson.

The classification of attainment in this study was based on teacher-reports against the norm of 'age related expectation', giving a trichotomised variable (below, at, or above). In their study, Jäger *et al.* (2015) used a median split to create a dichotomised variable of 'low' or 'high' achieving. This may have affected the pattern of results. Additionally, the effects were found for updating, but not for shifting or inhibition (Jäger *et al.*, 2015). Thus, it may be that the effect of attainment on behavioural engagement differs from the effect of attainment on updating.

As children's progress throughout their school career is based on normative assessments, and underachievement remains a focus for school leaders, the finding that those achieving below age-related expectation profit the most from PA participation is encouraging. It suggests that those pupils in need of additional support for their learning can receive a boost to their behavioural engagement by participation in PA during the school day. This warrants further research into the effects of PA for underachieving pupils.

### 5.4.3 Sports Participation

The typical weekly duration of sports participation was used in this study as a proxy of fitness. Once more using a median split, Jäger *et al.* (2015) identified effects on updating only for participants in the 'higher fit' category. In this study, self-reported sports participation was dichotomised as 'below' or

'at or above' sample mean minutes per week. Very small differences in effect sizes were identified, of which only one made the difference between a 'no effect' for those who were less active, and a small effect for those who were more active (medium intensity-low complexity). Overall, those in the less active group only benefited from the high intensity-high complexity condition, whereas those in the more active group profited from that condition as well as the medium intensity-low complexity condition. No effects beyond d=0.2 were identified for either group in the other four conditions.

The discordance between the present study and the work by Jäger *et al.* (2015) may be due to the use of a proxy measures rather than outcomes of fitness testing in my study. Due to time and budget restraints no fitness testing could be carried out, and proxy measures were the only available way of assessing PA habits. Fitness has also been found to moderate the response to acute PA in a study by Chang *et al.* (2014). They found that moderate fitness optimised participants' response to acute PA, which may not be identified when the measure of fitness is dichotomised. For future research into the effects of acute PA it is important to ensure fitness is also assessed. To sum up, those who spend more time taking part in organised sports showed a marginally more positive response to PA, but without a clear pattern to the directionality of this relationship based on the intensity and/or complexity of the PA.

Organised sports participation is not the only measure of PA, and some children may engage in lots of vigorous free play but not take part in organised sports. For future research, it would be interesting to also investigate the effects of interactions with free-living MVPA relative to the government recommendation of 60 minutes of MVPA per day.

### 5.4.4 Gender

Differential effects of PA on on-task behaviour were expected on the basis of participant gender, and this hypothesis was confirmed. However, the directionality of the gender effect was opposite to expected, with no effects greater than d=0.2 found for boys. Girls on the other hand responded positively to three of the six intervention conditions. Interestingly, the pattern of effect sizes was the

same for boys and girls (inverted-U for low complexity, and U-relationship for high complexity), but effect sizes for girls exceeded those for boys, so that three conditions exceeded d=0.2 for girls and none did for boys.

Boys were expected to respond more favourably to PA, as boys have been reported to spend less time on-task in lessons, and those with the lowest levels of on-task behaviour have been highlighted as benefiting the most from PA interventions. For example, Ma *et al.* (2014) found that passive off-task and motor off-task behaviours were reduced only for boys after PA breaks. Yet, Wilson *et al.* (2016) found no effects of a PA intervention on on-task behaviour or on the results of a computerised psychomotor vigilance task in an all-boys sample. Gender-based differences in the control variables did not provide an explanation for this difference in effect from the existing literature, and it would be interesting to replicate this aspect of the study with a larger sample.

In conclusion, differential effects for PA intensity and complexity were found based on individual differences, with the strongest effects for those with the lowest levels of on-task behaviour, academic achievement below age-related expectation, and females. As the sample in this study was quite small, further sub-divisions of the sample lead to groups with low numbers of participants and results must be interpreted with caution. Further research into the individual differences in behavioural responses to PA is needed with adequate sample sizes.

## 5.5 Research Question 5

As it has been found that PA positively influences emotion states, and emotion states are linked to task-focus and behaviour, it was hypothesised that the relationship between PA participation during PE lessons and subsequent on-task behaviour would be mediated by emotion states. It was found that, although medium intensity PE conditions increased participants' experiences of negative emotions both at the end of PE and in the following classroom lesson, and negative emotions in turn predicted decreased on-task behaviour, these effects did not combine into a significant mediation effect.

Additionally, mediation paths from PE enjoyment via PE lesson emotions and post-PE emotions to post-PE on-task behaviour were non-significant, indicating that despite PE enjoyment affecting the emotional state of participants at the end of PE and in the subsequent classroom lesson, this change did not significantly influence on-task behaviour. Thus, although an effect of PE lessons on emotion states was confirmed, and an effect of emotion states on post-PE on-task behaviour was found for negative emotions, the hypothesis of mediation by emotion states in the relationship between PA and on-task behaviour was rejected.

Despite strong evidence in the literature that acute PA positively influences emotions states (e.g. Chan *et al.*, 2018; Dunton *et al.*, 2014), and that emotion states are related to cognitive function and task-related behaviour (e.g. Pekrun, 2006; Meinhardt and Pekrun, 2003), only one published study was identified which investigated affect as a mediator in the relationship between simple and cognitively challenging PA, and measures of cognition. This study found that positive affect mediated the relationship between cognitively engaging breaks and focussed attention, but no direct or mediated effects were found for PA (Schmidt *et al.*, 2016).

In this study, positive effects were found of PE lesson enjoyment on positive emotions, and negative effects on negative emotions. However, PE enjoyment was not predicted by any particular type of PE lesson content. Negative emotions were also directly influenced by MPA, both with and without complexity. No direct effect of any PE intervention condition was found for positive emotions. This was unexpected, as the literature overwhelmingly indicates positive effects of PA participation on mood and emotion states (e.g. Chan *et al.*, 2018).

The discrepancies between the findings of this study and those reported in the literature could stem from a variety of sources. Some of these are rooted in the data collection, whereas others relate to the analytic procedures. With regard to the data collection, the aim was to track emotions throughout the afternoon during which the intervention took place. This involved repeated sampling and, to reduce the burden on participants as well as minimise disruption to their learning,

questionnaires were kept as short as possible. Two issues may stem from this approach; the repetitive nature of the sampling, and the number of items included in each self-report.

Firstly, the repeated reporting may have been too frequent, and not have allowed for the time period over which emotions naturally fluctuate to elapse between reports. Additionally, the frequency of reporting may have reduced interest and induced boredom or frustration in the participants, which may have affected the accuracy of their reporting. Secondly, the brevity of the self-report, with only eight emotion items included, may have caused difficulty for the analysis, which is scrutinised in more detail below. Yet, asking children in the target age group to report on eight different emotions, is already requiring a great deal of emotional awareness.

Regarding the data analysis, the data from the emotion reports was heavily skewed, which made the analysis more challenging. The heavy skew also indicates a degree of stability in the data, and a lesser extent of emotions being influenced by external factors. Alternatively, it could be the case that, as the sample consisted of young children, they were unable to identify with great precision the nature of the emotions they were experiencing at a given time. Another challenge in the analysis was related to the minimal number of emotion items. Fitting the factor structure was not a straightforward procedure, and eventually a two-factor structure was fitted rather than the hypothesised four-factor structure. This meant that the factors presenting the best fit included both activating and de-activating emotions both within the positive factor and within the negative factor. As it has been found that activation and de-activating emotions of the same valence (positive or negative) can have opposing effects of task-focus and attention (Pekrun, 2006; Pekrun *et al.*, 2002), this may have affected the subsequent findings on the relationships between emotion states and on-task behaviour. In fact, this may explain the lack of an influence identified for positive emotions on on-task behaviour throughout, as well as the absence of an effect of either positive or negative emotions on on-task behaviour in the pre-PE lessons.

Owing to the small sample size, the degrees of freedom in the SEM was limited. This means that to fit all the relationships between PE lesson intensity and complexity, PE lesson enjoyment, emotion

factors, and behavioural engagement, emotion reports had to be aggregated at the questionnaire level rather than be modelled as a factor structure throughout the levels. This may have affected the results from these models. Another consequence of the low number of degrees of freedom was that the same control variables identified as important in the MLM could not be included in the SEM. In the MLM, including the control variables brought out stronger effects for the predictor variables. Hence, it is possible that modelling limitation prevented some of the mediation effects of emotion states to be identified.

## 5.6 Implications for Practice

PE continues to suffer a poor reputation among teachers in UK primary schools. Often-cited barriers to teaching PE for primary school generalist teachers are the low status of the subject in the development of the child, a lack of curriculum accountability for PE at the school level coupled with pressure on performance in 'core' subjects of English, Maths, and Science (leading to teachers worrying about the time spent away from the classroom whilst delivering PE), the need for burdensome equipment preparation or a lack of available equipment or facilities, and a lack of teachers' confidence and skills in delivering high quality and safe PE lessons due to poor teacher training and continued professional development (e.g. Tsangaridou, 2017; Morgan and Hansen, 2008a; Harris *et al.*, 2012). Some of these barriers are related to the individual teacher, whereas others exist at the level of the school or policies. In this section the implications of this study for individual teachers are discussed. Implications for policy are considered in section 5.7.

The key message to teachers from this study is that careful planning of the content of PE lessons can maximise the benefits to pupils' learning in the subsequent lesson. The results from this study provide vital evidence that the benefits of PE are not limited to the sports field, and that PE has a positive impact on classroom learning. By demonstrating that PE lesson content can be used to help improve pupils' task-related behaviour, especially for those who are the least on-task and provide the most disruption in class, this study makes explicit to teachers one of the possible reasons why so many

studies have found that increasing the time spent on PA and PE does not have a detrimental effect on academic performance. Teachers want pupils to acquire knowledge and skills during their lessons, and for them to do so without disruption. That is where the main impact of this study lies: it provides primary school teachers with a new tool to tackle inattentiveness and disruption in their classrooms. On-task behaviour is a tangible thing; the impact of this study was visible to the teachers in their classrooms, and it was immediate.

To harness the potential of PE lessons for on-task behaviour, there are concrete changes teachers can make to their practice based on the results from this study. Firstly, by demonstrating to teachers how a range of simple PE lessons with low equipment requirements can benefit subsequent classroom behaviour, the results from this study should make teachers think twice about skipping PE lessons in favour of English, maths, or science lessons. Secondly, teachers can try to avoid scheduling PE as the last lesson before home time, as the beneficial effects for on-task behaviour will be lost. Thirdly, if there are particularly vital lessons or assessments to be completed, these can be scheduled for after a PE lesson. Especially as those pupils who are most off-task were found to benefit most from the intervention, and they significantly benefited from each condition in the intervention, this gives teachers a new strategy to assist these pupils to remain on-task and boost their potential for learning. Meanwhile, this reduces disruption to the whole class caused by the off-task behaviour of a few pupils. As outlined in my conceptual model displayed in Figure 7 (p.52), on-task behaviour is one of the pathways by which PA contributes to academic achievement.

By making the benefits of PE to classroom learning more explicit, this study also contributes to the growing evidence that PE and PA have an important role to play in the cognitive development of children. More teachers need to be brought on-board with this viewpoint in order to change the current and long-held perception that PE is of lower educational value to other subjects. It is time to bring PE out of the margins it has been pushed into for far too long.

For primary school teachers another barrier to teaching PE is the subject's high dependency on equipment, and the accompanying burdensome equipment set-up. This study has shown that

effective PE lessons do not require extensive or difficult to operate equipment. All lessons from this intervention can be taught with basic equipment (cones, tennis balls, footballs, bibs), which is widely found in UK primary schools. This was purposely done to ensure the benefits of this research would be available to as many teachers and children as possible.

This study provides knowledge to teachers, and helps them to understand the important role PE plays in the development of their pupils, both physically and cognitively. However, a wider change to PE in primary schools is needed to maximise the benefits to all pupils. Changes are required at the level of policy as well as practice, which will be discussed in the next section.

## 5.7 Implications for Policy

The results of this study provide tools for teachers to make deliberate choices about PE lesson content (section 5.6), and to deliver high quality PE lessons that satisfy the National Curriculum as well as improve on-task behaviour in subsequent classroom lessons. However, the individual teachers cannot do all of this on their own; to make the benefits available to all teachers and all pupils, teachers need to be supported at the level of the institution(s) they work for. To facilitate changes to practice, changes are needed to policies on teacher education, and the regulations for primary school timetabling and time allocation to PA.

Considering the potential for political and policy change in light of the general election in December 2019 and Brexit, the UK is currently in an interesting position politically. However, there still is a lack of attention to research evidence in the educational policy making sector. Over the past decade, funding for education has declined, and many schools have been forced to sell off the land on which their playing fields were located. Facilities for sports and PA have been reduced, as has the time allocated to PE, break times, and lunch. All of this has happened despite increasing research evidence on the importance of PA for physical health (Cumming and Riddoch, 2008), mental health (Chan *et al.*, 2018; Biddle and Asare, 2011), and cognitive performance (Álvarez-Bueno *et al.*, 2017).

Firstly, changes are required to policies on teacher training. The results from this study contribute to the overall body of literature in support of PA in school. Thereby, this study adds weight to the argument for those in charge of teacher training to allocate more time and resources to preparing both future and current teachers to teach PE or physically active lessons, and incorporate PA in the school day. The knowledge and understanding about the all-round benefits of PA needs to be disseminated to all teachers, through initial teacher training and continued professional development. This requires a major adaptation to the skills set of primary school teachers, many of whom have received no training in PE and therefore lack the skills and confidence to deliver this subject (Iserbyt *et al.*, 2017; Harris *et al.*, 2012). Increasing the time and attention devoted to PE and other forms of structured PA (such as physically active lessons and short classroom PA breaks) during teaching training can make a huge difference to teachers and their practice (Harris, Cale and Musson, 2011). Giving teachers the knowledge to choose the right content for their pupils, along with the confidence and skills to deliver this content, is key. This study contributes to the knowledge element that can be shared with teachers, but they also need and deserve appropriate training on how to apply this to their daily practice. As determined by Lubans *et al.* (2017) among others, delivery of the content as well as the content itself is vital. These kinds of changes to practice will not happen by themselves and need to be actioned through teacher training policy. A strong and wide literature base is required to force these changes to policy, and that is where this study makes a contribution.

Secondly, changes to time allocation and timetabling in primary schools are essential. Not only do we need a minimum requirement of weekly time spent on PA, but also an enforced spread of PA across the week so that PA becomes a habit, and something that children do daily. The acute effects of PA are of limited duration, and children should get these benefits throughout the week, on every day. With the current focus in schools remaining on standardised tests and league tables, school leaders are unlikely to enforce daily PA time, despite overwhelming research evidence that additional time spent in PA does not negatively affect academic achievement – and in many cases improves it (e.g. De Greeff *et al.*, 2018; Verburgh *et al.*, 2014; Lees and Hopkins, 2013). As my proposed model highlights,

based on existing research and literature, PA is a way to contribute to academic achievement. Yet, school leaders continue to believe it detracts from achievement. For this reason, it is important to have these changes applied at the highest level of policy making, rather than allowing schools to make their own decision on this matter.

The lack of attention for research evidence in educational policy making, and for school-based PA in particular, has been highlighted again in the 2019 election manifestos of the four main political parties. The manifesto of the current governing party for the 2019 election fares the best on primary school PE and PA. It promises an investment of £14billion for schools, which translates to an additional £4,000 a year for each primary school pupil (Conservative and Unionist Party, 2019). One of the main drivers behind this additional funding is their aim to raise standards, and the manifesto acknowledges that "students learn best (…) in an environment where there is no disruption" (ibid.: 13). Additionally, their manifesto includes a specific pledge to support primary school PE teaching, to "ensure that it is being properly delivered" (ibid.: 13), and to maximise the use of sports facilities to "promote physical literacy and competitive sports" (ibid.: 13).

Likewise, the one of the main parties' manifestos includes a pledge to increase spending on education by at least £4billion per year (Green Party, 2019). They aim to make classrooms a more pleasant environment, and want all children to spend at least the equivalent of half a day in sports (ibid.). Although they don't specifically mention classrooms free of disruption, but rather a 'pleasant environment', theirs is the only manifesto to name a minimum duration for sport. In fact, by choosing the word 'sport' rather than PA they have implicitly excluded the informal PA children get during break and lunchtime. As not all children choose to be active during breaks, this choice of wording has important implications for ensuring that this time allocation is spent as actual active time. An important difference from the current governing party's pledges though is that there is no focus on the quality of the active time. This indicates that there is still a great need to get politicians and policy makers to pay more attention to research evidence.

On the other hand, both other main parties make no specific mention of school-based PA in their manifestos at all. One does stress the need to shift the focus of schools away from the core subjects of English, Maths, and Science, onto a wider range of subjects, and it wants to abolish high stakes testing in primary schools (Labour Party, 2019). Great emphasis is placed on arts and creative subjects to achieve this, but the manifesto makes no pledges on PA in schools (ibid.). Nor does it mention in the part of the manifesto dedicated to education anything about disruption to learning, behaviour in schools, or a focus on learning (ibid.). Its 'Social Transformation Fund' of £150billion is pledged to schools, hospitals, social care, and council housing, with no details of the breakdown of this amount of money between the four sectors (Labour Party, 2019). Altogether this demonstrates a severe disjoint from several serious issues in education, and a lack of incorporation of research evidence in policy making.

Another party pledges to introduce a 'Curriculum for Life' which ensures providing children with a well-rounded education, rather than focusing solely on the core subjects (Liberal Democrats, 2019). The Curriculum for Life would include "Personal, Social and Health Education, financial literacy, environmental awareness, first aid and emergency lifesaving skills, mental health education, citizenship and age-appropriate Relationships and Sex Education (RSE). Teaching about sexual consent, LGBT+ relationships, and issues surrounding explicit images and content will be included in RSE." (ibid.: 33). No mention of PA, PE, or sport in the context of schools and education is made in their manifesto. In fact, the sole mention of sports is in a pledge to protect sports funding through the National Lottery in relation to schools and education (ibid.). The importance of PA is mentioned only in recognition of its importance for physical health (in relation to medical care and the NHS), and the need to support active play opportunities for children (ibid.). No specific strategies or plans are provided for how this will be achieved.

Addressing important points raised by the currently governing party's manifesto, this study provides evidence that PE lesson content does matter, and what teachers choose to deliver makes a real difference to their pupils' learning. Not only does PE teach physical literacy and contribute to the

physical health of pupils, it can also assist in tackling the earlier point of creating positive environments for learning. If the content of the lesson is chosen with care and deliberation, based on the teacher's knowledge and understanding of the benefits of PA, PE can help to provide an environment with reduced disruption and in which all pupils are able to optimise their learning. We need politicians and policy makers on board to ensure teachers are provided with the support they need to make the most of the knowledge we have uncovered through research over the past decades. This study contributes to the growing school-based and practically situated evidence base required to get this point firmly placed on the political agenda.

## 5.8 Strengths and Limitations

### 5.8.1 Strengths

This study has several strengths. First of all, the naturalistic setting of the research has lent it strong ecological validity. The research questions were informed by research findings from laboratory study and studies investigating measures of cognition in psychometric tasks. By conducting this research in the classroom and selecting an outcome measure which is directly visible and valuable to teachers, this research powerfully relates to daily life in primary schools. Another aspect of this relatability is the choice to investigate PE lessons, an already occurring period of PA in schools, rather than putting further demands on teachers to dedicate more time to PA. Although research evidence overwhelmingly suggests there is no evidence of detrimental effects of reallocating classroom time to PA, teachers remain reluctant to take additional time away from academic instruction. The premise of this study was to take PE lessons and see if the content of the lessons can be optimised for subsequent learning, whilst still fitting within the requirements of the curriculum, and with the use of only basic PE equipment such as can be found in practically every UK primary school.

Secondly, the study was designed to systematically manipulate both the intensity and complexity of PA in a way that has not been done before. Four-way interventions of low and high PA demands in

combination with low and high cognitive demands have been reported, but a six-way comparison with three levels of intensity across both low and high complexity is a novel approach. Additionally, applying a within-person design to this approach has enabled the investigation of relative differences in the response to specific manipulations of PA, rather than comparing the occurrence of PA to a no-PA control condition. Momentary experience sampling in the form of self-reports was completed throughout every intervention condition and provided control variables to account for fluctuations in contextual variables.

Moreover, direct observation was used in this study to record on-task behaviour. This has been proposed as the preferred method to assess classroom behaviour, as being present in the classroom allows the rater to be fully immersed in the context in which the observations take place. And, finally, advanced quantitative analysis strategies allowed for the modelling of fluctuations in behaviour as a combination of moment-to-moment effects (time-points), week-to-week effects (intervention condition), and person-to-person effects (individual differences).

### 5.8.2 Limitations

Some limitations must be noted, however. Conducting research in real-world settings, and especially in schools, reduces the level of control the researcher has and introduces numerous contextual factors and influences. Effect sizes tend to be smaller in real-life settings, and it may be that the low incidence of effect sizes achieving practical significance above d=0.2 was related to this choice of setting. Moreover, being restricted by the timings of the school day meant that observation periods were limited to 25 minutes, and potential benefits of PA to attention might have been missed if they occurred outside of this time frame. For example, Schmidt *et al.* (2015a) found effects of cognitively engaging PA on d2 test performance not immediately after cessation of exercise, but 90 minutes later.

The decision to use direct observation of behaviour meant that data collection was time consuming and intense. With limited time available, this resulted in a limited sample size and a reduction to the

collection of certain background variables, such as measures of fitness. The limited sample size restricted the complexity of some of the SEM, which may have hidden potential effects of PA and emotion states on on-task behaviour.

The measurement of PA intensity in this study was affected by the level of complexity, due to the accelerometers being worn on the wrist and high-complexity lessons at the low and medium intensity levels involving coordinative exercise made up of object manipulation. Although this could have been avoided by selecting a different site for children to wear the accelerometer, the wrist has been indicated as the ideal site for wear in free-living circumstances owing to enhanced comfort and thus wear compliance. As free-living activity was also of interest in this study, a trade-off of precision of measurement during PE and wear compliance in free-living data collection was inevitable.

Lastly, the level of complexity in the PE lessons was not quantified in this study, other than 'low' or 'high'. This may have allowed for varying levels of complexity in the different high-complexity conditions, which could affect the findings. Recently, work by Schmidt *et al.* (2019) indicates the existence of a dose-response relationship of the cognitive load of PA beyond the presence or absence of cognitive challenge, and dependent on the individual's need for cognitive challenge.

### 5.8.3 Recommendations for Future Research

On the basis of the findings of this study, and taking into account the limitations, suggestions for future research include replication of the current study with an increased sample size. This would allow for greater detail in the investigation of mediation effects, as well as the examination of individual differences without loss of statistical power due to subdivision of the sample. It may also permit the inclusion of participants who needed to be excluded from this study on medical grounds, despite being able to take part in PE lessons. Research focused on particular sub-groups is still scarce, with inconclusive study results reported. This would be a valuable direction to explore further, and would help to uncover some of the specific mechanisms by which PA can contribute to children's cognitive development.

It would be beneficial to include additional measures in a future design. Assessing participants' fitness would allow for an additional control variable in the statistical modelling of effects of aerobic intensity, as fitness level has been identified as a potential moderator. The same applies to gross motor skills in the analyses of complexity – a participants' gross motor skill level influences the degree of challenge presented in the PE lesson and would be important to include as a mediator or control variable. The inclusion of a test of motor skills would help to quantify the level of cognitive challenge. This could prevent or purposely induce the occurrence of cognitive overload to test for effects, so that a more informative conclusion can be drawn about the effect of exercise complexity as well as the interaction of exercise intensity and complexity.

Another possibility is a more in-depth investigation of the relationships between school-based PA, mood, and emotions on the one hand, and task focus on the other. Literature on this topic in primary school children is sparse, possibly related to limitations in many children's emotional literacy at this age. Development of a new research instrument aimed at repeated momentary assessments in the primary school age range may be required in order to carry out such a study.

Finally, a potential focus for future research would be to further investigate the variation in results achieved in studies with different outcome measures, such as behaviour observations versus cognitive function tests. Examining the relationship between PA induced changes to cognitive function test results and classroom behaviour could reveal the reason for the discrepancy in results. Additionally, for teachers this would lend more practical, classroom-based relevance to the body of literature on physical activity and cognitive functions.

# 6 Conclusion

These findings suggest that participation in PE lessons can improve children's on-task behaviour in the classroom, in particular for those students with low levels of on-task behaviour. PE lessons involving high complexity and high intensity appear to have the greatest benefits for students' behaviour in the classroom.

Hypothesis one and two, regarding a main effect of PE lesson complexity and intensity on subsequent on-task behaviour were rejected. Hypothesis three, that complexity and intensity interact to affect on-task behaviour was accepted, as was hypothesis four, that individual differences influence the effect of PE lesson content on on-task behaviour. However, the direction of the hypothesised relationships was not as expected with the exception of low pre-PE on-task behaviour, for which the hypothesis was fully accepted. Finally, hypothesis five, that the effect of PE lesson complexity and intensity on subsequent on-task behaviour is mediated by emotion states, was rejected.

# References


Aadland, K.N., Moe, V.F., Aadland, E., Anderssen, S.A., Resaland, G.K. and Ommundsen, Y. (2017) 'Relationships between physical activity, sedentary time, aerobic fitness, motor skills and executive function and academic performance in children'. *Mental Health and Physical Activity* 12 pp.10-18. DOI: 10.1016/j.mhpa.2017.01.001.

Activinsights. (2012) GENEActiv Instructions.

Álvarez-Bueno, C., Pesce, C., Cavero-Redondo, I., Sánchez-López, M., Garrido-Miguel, M. and Martínez-Vizcaíno, V. (2017) 'Academic Achievement and Physical Activity: A Meta-analysis'. *Pediatrics* 140(6) pp.e20171498. DOI: 10.1542/peds.2017-1498.

Alves, C.R.R., Tessaro, V.H., Teixeira, L.A.C., Murakava, K., Roschel, H., Gualano, B. and Takito, M.Y. (2014) 'Influence of Acute High-Intensity Aerobic Interval Exercise Bout on Selective Attention and Short-Term Memory Tasks'. *Perceptual and Motor Skills* 118(1) pp.63-72. DOI: 10.2466/22.06.PMS.118k10w4.

Andrew, K., Richards, R., Gaudreault, K.L. and Woods, A.M. (2016) 'Understanding physical educators' perceptions of mattering'. *European Physical Education Review* 23(1) pp.73-90. DOI: 10.1177/1356336X16637320.

Armstrong, N. and Fawkner, S.G. (2008) Exercise metabolism. In: Armstrong, N. and Van Mechelen, W. (eds) *Paediatric Exercise Science & Medicine.* 2nd ed. Oxford: Oxford University Press, pp.213-226.

Bailey, R. (2006) 'Physical Education and Sport in Schools: A Review of Benefits and Outcomes'. *Journal of School Health* 76(8) pp.397-401. DOI: 10.1111/j.1746-1561.2006.00132.x.

Barros, R.M., Silver, E.J. and Stein, R.E. (2009) 'School recess and group classroom behavior'. *Pediatrics* 123(2) pp.431-436. DOI: 10.1542/peds.2007-2825.

Bartholomew, J.B., Golaszewski, N.M., Jowers, E., Korinek, E., Roberts, G., Fall, A. and Vaughn, S. (2018) 'Active learning improves on-task behaviors in 4th grade children'. *Preventive medicine* 111 pp.49-54. DOI: 10.1016/j.ypmed.2018.02.023.

Berchicci, M., Pontifex, M.B., Drolette, E.S., Pesce, C., Hillman, C.H. and Di Russo, F. (2015) 'From Cognitive Motor Preparation to Visual Processing: The Benefits of Childhood Fitness to Brain Health'. *Neuroscience* (298) pp.211-219. DOI: 10.1016/j.neuroscience.2015.04.028.

Berlyne, D.E. (1966) 'Curiosity and Exploration'. *Science, New Series* 153(3731) pp.25-33.

Best, J.R. (2012) 'Exergaming immediately enhances children's executive function'. *Developmental psychology* 48(5) pp.1501-1510. DOI: 10.1037/a0026648.

Biddle, S.J.H. and Asare, M. (2011) 'Physical activity and mental health in children and adolescents: a review of reviews'. *British Journal of Sports Medicine* 45(11) pp.886. DOI: 10.1136/bjsports-2011-090185.

Bloom, H.S. (2012) The Core Analytics of Randomized Experiments for Social Research. In: Alasuutari, P., Bickman, L. and Brannen, J. (eds) *The SAGE Handbook of Social Research Methods.* London: SAGE Publications Ltd, pp.115-133.

Borenstein, M., Hedges, L.V., Higgins, J.P.T. and Rothstein, H.R. (2009) Converting Among Effect Sizes. In: Borenstein, M. (ed) *Introduction to meta-analysis.* Chichester: John Wiley & Sons.

Botvinick, M.M., Braver, T.S., Barch, D.M., Carter, C.S. and Cohen, J.D. (2001) 'Conflict Monitoring and Cognitive Control'. *Psychological Review* 108(3) pp.624-652. DOI: 10.1037//0033-295X.108.3.624.

Bowers, J.S. (2016) 'The practical and principled problems with educational neuroscience'. *Psychological Review* 123(5) pp.600-612. DOI: 10.1037/rev0000025.

Broadbent, D.E. (1953) 'Classical conditioning and human watch-keeping'. *Psychological Review* 60(5) pp.331-339.

Brockett, A.T., LaMarca, E.A. and Gould, E. (2015) 'Physical exercise enhances cognitive flexibility as well as astrocytic and synaptic markers in the medial prefrontal cortex'. *PLoS One* 10(5) pp.e0124859. DOI: 10.1371/journal.pone.0124859.

Browne, M.W. and Cudeck, R. (1992) 'Alternative Ways of Assessing Model Fit'. *Sociological Methods & Research* 21(2) pp.230-258. DOI: 10.1177/0049124192021002005.

Bruer, J.T. (1997) 'Education and the Brain: A Bridge Too Far'. *Educational Researcher* 26(8) pp.4-16. DOI: 10.2307/1176301.

Buck, S.M., Hillman, C.H. and Castelli, D.M. (2008) 'The relation of aerobic fitness to stroop task performance in preadolescent children'. *Medicine and Science in Sports and Exercise* 40(1) pp.166-172. DOI: 10.1249/mss.0b013e318159b035.

Budde, H., Voelcker-Rehage, C., Pietrabyk-Kendziorra, S., Ribeiro, P. and Tidow, G. (2008) 'Acute coordinative exercise improves attentional performance in adolescents'. *Neuroscience Letters* 441(2) pp.219-223. DOI: 10.1016/j.neulet.2008.06.024.

Bugler, M., McGeown, S.P. and St Clair-Thompson, H. (2015) 'Gender differences in adolescents' academic motivation and classroom behaviour'. *Educational Psychology* 35(5) pp.541-556. DOI: 10.1080/01443410.2013.849325.

Byrne, B.M. (2009) *Structural equation modeling with AMOS: basic concepts, applications, and programming,* New York: Routledge.

Byun, K., Hyodo, K., Suwabe, K., Ochi, G., Sakairi, Y., Kato, M., Dan, I. and Soya, H. (2014) 'Positive effect of acute mild exercise on executive function via arousal-related prefrontal activations: An fNIRS study'. *NeuroImage* 98 pp.336-345. DOI: 10.1016/j.neuroimage.2014.04.067.

Capaldi, C.A., Dopko, R.L. and Zelenski, J.M. (2014) 'The relationship between nature connectedness and happiness: a meta-analysis'. *Frontiers in Psychology* 5 pp.976. DOI: 10.3389/fpsyg.2014.00976.

Carlson, J.A., Engelberg, J.K., Cain, K.L., Conway, T.L., Mignano, A.M., Bonilla, E.A., Geremia, C. and Sallis, J.F. (2015) 'Implementing classroom physical activity breaks: Associations with student physical activity and classroom behavior'. *Preventive medicine* 81 pp.67-72. DOI: 10.1016/j.ypmed.2015.08.006.

Castelli, D.M., Hillman, C.H., Buck, S.M. and Erwin, H.E. (2007) 'Physical Fitness and Academic Achievement in Third- and Fifth-Grade Students'. *Journal of Sport & Exercise Psychology* 29(2) pp.239-252. DOI: 10.1123/jsep.29.2.239.

Chaddock, L., Pontifex, M.B., Hillman, C.H. and Kramer, A.F. (2011) 'A review of the relation of aerobic fitness and physical activity to brain structure and function in children'. *Journal of the International Neuropsychological Society* 17(6) pp.975-985. DOI: 10.1017/S1355617711000567.

Chan, J.S.Y., Liu, G., Liang, D., Deng, K., Wu, J. and Yan, J.H. (2018) 'Therapeutic Benefits of Physical Activity for Mood: A Systematic Review on the Effects of Exercise Intensity, Duration, and Modality'. *The Journal of Psychology* (Special Issue) pp.1-24. DOI: 10.1080/00223980.2018.1470487.

Chang, Y.-K., Chi, L., Etnier, J.L., Wang, C.-C., Chu, C.-H. and Zhou, C. (2014) 'Effect of acute aerobic exercise on cognitive performance: Role of cardiovascular fitness'. *Psychology of Sport and Exercise* 15(5) pp.464-470. DOI: 10.1016/j.psychsport.2014.04.007.

Chang, Y.K., Chu, C.H., Wang, C.C., Wang, Y.C., Song, T.F., Tsai, C.L. and Etnier, J.L. (2015) 'Dose-response relation between exercise duration and cognition'. *Medicine & Science in Sports & Exercise* 47(1) pp.159-165. DOI: 10.1249/MSS.0000000000000383.

Chang, Y.K., Labban, J.D., Gapin, J.I. and Etnier, J.L. (2012) 'The effects of acute exercise on cognitive performance: A meta-analysis'. *Brain Research* 1453 pp.87-101. DOI: 10.1016/j.brainres.2012.02.068.

Charlton, C., Rasbash, J., Browne, W.J., Healy, M. and Cameron, B. (2019) MLwiN Version 3.04. Available at: http://www.bristol.ac.uk/cmm/software/mlwin/download/ (Downloaded: 09.07.2019).

Chekroud, S.R., Gueorguieva, R., Zheutlin, A.B., Paulus, M., Krumholz, H.M., Krystal, J.H. and Chekroud, A.M. (2018) 'Association between physical exercise and mental health in 1·2 million individuals in the USA between 2011 and 2015: a cross-sectional study'. *The Lancet Psychiatry* 5(9) pp.739-746. DOI: 10.1016/S2215-0366(18)30227-X.

Chen, A.-G., Yan, J., Yin, H.-C., Pan, C.-Y. and Chang, Y.-K. (2014) 'Effects of acute aerobic exercise on multiple aspects of executive function in preadolescent children'. *Psychology of Sport & Exercise* 15(6) pp.627-636. DOI: 10.1016/j.psychsport.2014.06.004.

Chief Medical Officers. (2011) *Physical Activity Guidelines for Children and Young People*. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/213739/dh_128144.pdf (Accessed: 14.11.2018).

Chomitz, V.R., Slining, M.M., McGowan, R.J., Mitchell, S.E., Dawson, G.F. and Hacker, K.A. (2009) 'Is there a relationship between physical fitness and academic achievment? Positive results from public school children in the Northeastern United States.'. *Journal of School Health* 79(1) pp.30-37. DOI: 10.1111/j.1746-1561.2008.00371.x.

Coe, D.P., Pivarnik, J.M., Womack, C.J., Reeves, M.J. and Malina, R.M. (2006) 'Effect of physical education and activity levels on academic achievement in children'. *Medicine and Science in Sports and Exercise* 38(8) pp.1515-1519. DOI: 10.1249/01.mss.0000227537.13175.1b.

Colman, A.M. (2015) *A dictionary of psychology*. Available at: https://www.oxfordreference.com/view/10.1093/acref/9780199657681.001.0001/acref-9780199657681 (Accessed: 07.09.2017).

Conservative and Unionist Party. (2019) The Conservative and Unionist Party Manifesto 2019. London: Paragon CC.

Cook, T.D. and Wong, V.C. (2012) Better Quasi-Experimental Practice. In: Alasuutari, P., Bickman, L. and Brannen, J. (eds) *The SAGE Handbook of Social Research Methods.* London: SAGE Publications Ltd, pp.134-165.

Coolican, H. (2009) *Research methods and statistics in psychology,* London: Hodder Education.

Covassin, T., Weiss, L., Powell, J. and Womack, C. (2007) 'Effects of a maximal exercise test on neurocognitive function'. *British Journal of Sports Medicine* 41(6) pp.370-374. DOI: 10.1136/bjsm.2006.032334.

Crews, D.J., Lochbaum, M.R. and Landers, D.M. (2004) 'Aerobic physical activity effects on psychological well-being in low-income hispanic children'. *Perceptual and Motor Skills* 98 pp.319-324. DOI: 10.2466/pms.98.1.319-324.

Cumming, S. and Riddoch, C. (2008) Physical activity, physical fitness, and health: current concepts. In: Armstrong, N. and Van Mechelen, W. (eds) *Paediatric Exercise Science and Medicine.* 2nd ed. Oxford: Oxford University Press.

Davidson, M.C. and Morrocco, R.T. (2000) 'Local infusion of scopoplamine into intraparietal cortex slows covert orienting in rhesus monkeys'. *Journal of Neuroscience* 83 pp.1536-1549. DOI: 0022-3077/00.

Davis, C.L., Tomporowski, P.D., Boyle, C.A., Waller, J.L., Miller, P.H., Naglieri, J.A. and Gregoski, M. (2007) 'Effects of aerobic exercise on overweight children's cognitive functioning: a randomized controlled trial'. *Research Quarterly for Exercise and Sport* 78(5) pp.510-519. DOI: 10.1080/02701367.2007.10599450.

De Greeff, J.W., Bosker, R.J., Oosterlaan, J., Visscher, C. and Hartman, E. (2018) 'Effects of physical activity on executive functions, attention and academic performance in preadolescent children: a meta-analysis'. *Journal of Science and Medicine in Sport* 21(5) pp.501-507. DOI: 10.1016/j.jsams.2017.09.595.

De Jong, K., Moerbeek, M. and van der Leeden, R. (2010) 'A priori power analysis in longitudinal three-level multilevel models: an example with therapist effects'. *Psychotherapy Research* 20(3) pp.273-284. DOI: 10.1080/10503300903376320.

Department for Education. (2013) *The national curriculum in England. Key stages 1 and 2 framework document.* Available at: https://www.gov.uk/government/publications/national-curriculum-in-england-primary-curriculum (Accessed: 01.06.2017).

Department for Education. (2019) *Healthy Schools Rating Scheme.* Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/814914/Healthy_schools_rating_scheme.pdf (Accessed: 12.08.2019).

Department for Education and Education Funding Agency. (2016a) *PE and sport premium conditions of grant 2016 to 2017: academies and free schools.* Available at: https://www.gov.uk/government/publications/pe-and-sport-premium-funding-conditions-for-2016-to-2017/pe-and-sport-premium-conditions-of-grant-2016-to-2017-academies-and-free-schools (Accessed: 28.02.2017).

Department for Education and Education Funding Agency. (2016b) *PE and sport premium for primary schools.* Available at: https://www.gov.uk/guidance/pe-and-sport-premium-for-primary-schools (Accessed: 28.02.2017).

Donnelly, J.E., Hillman, C.H., Castelli, D.M., Etnier, J.L., Lee, S., Tomporowski, P., Lambourne, K. and Szabo-Reed, A.N. (2016) 'Physical Activity, Fitness, Cognitive Function, and Academic Achievement in Children: A Systematic Review'. *Medicine and Science in Sports and Exercise* 48(6) pp.1197-1222. DOI: 10.1249/MSS.0000000000000966.

Duncan, M.J., Wilson, S., Tallis, J. and Eyre, E. (2016) 'Validation of the Phillips et al. GENEActiv accelerometer wrist cut-points in children aged 5-8 years old'. *European Journal of Pediatrics* 175(12) pp.2019-2021. DOI: 10.1007/s00431-016-2795-6.

Dunton, G.F., Huh, J., Leventhal, A.M., Riggs, N., Hedeker, D., Spruijt-Metz, D. and Pentz, M.A. (2014) 'Momentary assessment of affect, physical feeling states, and physical activity in children'. *Health Psychology* 33(3) pp.255. DOI: 10.1037/a0032640.

Egger, F., Benzing, V., Conzelmann, A. and Schmidt, M. (2019) 'Boost your brain, while having a break! The effects of long-term cognitively engaging physical activity breaks on children's executive functions and academic achievement'. *PLoS One* 14(3) pp.e0212482. DOI: 10.1371/journal.pone.0212482.

Egger, F., Conzelmann, A. and Schmidt, M. (2018) 'The effect of acute cognitively engaging physical activity breaks on children's executive functions: Too much of a good thing?'. *Psychology of Sport & Exercise* 36(C) pp.178-186. DOI: 10.1016/j.psychsport.2018.02.014.

Eime, R.M., Young, J., Harvey, J., Charity, M. and Payne, W. (2013) 'A systematic review of the psychological and social benefits of participation in sport for children and adolescents: informing development of a conceptual model of health through sport'. *International Journal of Behavioral Nutrition and Physical Activity* 10(1). DOI: 10.1186/1479-5868-10-98.

Elliot, J., Holland, J. and Thomson, R. (2012) Longitudinal and Panel Studies. In: Alasuutari, P., Bickman, L. and Brannen, J. (eds) *The SAGE Handbook of Social Research Methods.* London: SAGE Publications Ltd, pp.228-248.

Erickson, K.I., Hillman, C.H. and Kramer, A.F. (2015) 'Physical activity, brain, and cognition'. *Current Opinion in Behavioral Sciences* 4 pp.27-32. DOI: https://doi.org/10.1016/j.cobeha.2015.01.005.

Esliger, D.W., Rowlands, A.V., Hurst, T.L., Catt, M., Murray, P. and Eston, R.G. (2011) 'Validation of the GENEA Accelerometer'. *Medicine & Science in Sports & Exercise* 43(6) pp.1085-1093. DOI: 10.1249/MSS.0b013e31820513be.

Fairclough, S.J., Noonan, R., Rowlands, A.V., Van Hees, V., Knowles, Z. and Boddy, L.M. (2016) 'Wear Compliance and Activity in Children Wearing Wrist- and Hip-Mounted Accelerometers'. *Medicine and Science in Sports and Exercise* 48(2) pp.245-253. DOI: 10.1249/MSS.0000000000000771.

Faul, F., Erdfelder, E., Lang, A.-G. and Buchner, A. (2017) G*power, version 3.0.1. Available at: https://download.cnet.com/developer/department-of-psychology (Downloaded: 08.10.2017).

Fawkner, S.G. and Armstrong, N. (2008) Oxygen uptake kinetics. In: Armstrong, N. and Van Mechelen, W. (eds) *Paediatric Exercise Science and Medicine.* 2nd ed. Oxford: Oxford University Press.

Fedewa, A.L. and Ahn, S. (2011) 'The effects of physical activity and physical fitness on children's achievement and cognitive outcomes: a meta-analysis'. *Research Quarterly for Exercise and Sport* 82(3) pp.521-535. DOI: 10.1080/02701367.2011.10599785.

Field, A. (2013) *Discovering Statistics Using IBM SPSS Statistics,* London: SAGE.

Field, T., Diego, M. and Sanders, C.E. (2001) 'Exercise is positively related to adolescents' relationships and academics'. *Adolescence* 36(141) pp.105-110.

Frankmann, J.P. and Adams, J.A. (1962) 'Theories of vigilance'. *Psychological Bulletin* 59(4) pp.257-272.

Freak, A. and Miller, J. (2017) 'Magnifying pre-service generalist teachers' perceptions of preparedness to teach primary school physical education'. *Physical Education and Sport Pedagogy* 22(1) pp.51-70. DOI: 10.1080/17408989.2015.1112775.

Fredricks, J.A., Blumenfeld, P.C. and Paris, A.H. (2004) 'School Engagement: Potential of the Concept, State of the Evidence'. *Review of Educational Research* 74(1) pp.59-109. DOI: 10.3102/00346543074001059.

Frith, E., Sng, E. and Loprinzi, P.D. (2017) 'Randomized controlled trial evaluating the temporal effects of high-intensity exercise on learning, short-term and long-term memory, and prospective memory'. *European Journal of Neuroscience* 46(10) pp.2557-2564. DOI: 10.1111/ejn.13719.

Frühauf, A., Niedermeier, M., Elliott, L.R., Ledochowski, L., Marksteiner, J. and Kopp, M. (2016) 'Acute effects of outdoor physical activity on affect and psychological well-being in depressed patients – A preliminary study'. *Mental Health and Physical Activity* 10 pp.4-9. DOI: 10.1016/j.mhpa.2016.02.002.

Gallotta, M.C., Emerenziani, G.P., Franciosi, E., Meucci, M., Guidetti, L. and Baldari, C. (2015) 'Acute physical activity and delayed attention in primary school students'. *Scandinavian Journal of Medicine & Science in Sports* 25(3) pp.e331-e338. DOI: 10.1111/sms.12310.

Gauvin, L., Rejeski, W.J. and Norris, J.L. (1996) 'A naturalistic study of the impact of acute physical activity on feeling states and affect in women'. *Health Psychology* 15(5) pp.391. DOI: 10.1037/0278-6133.15.5.391.

Geiser, C. (2013) *Data analysis with Mplus,* New York: Guilford Press.

Godwin, K.E., Almeda, M., Petroccia, M., Baker, R.S. and Fisher, A.V. (2013) 'Classroom activities and off-task behavior in elementary school children', *35th Annual Conference of the Cognitive Science Society.* Austin, Tx, Cognitive Science Society. pp.2428-2433.

Godwin, K.E. and Fisher, A.V. (2011) 'Allocation of Attention in Classroom Environments: Consequences for Learning', *33rd Annual Conference of the Cognitive Science Society.* Austin, Tx, Cognitive Science Society. pp.2806-2811.

Goetz, T., Pekrun, R., Hall, N. and Haag, L. (2006) 'Academic emotions from a social-cognitive perspective: Antecedents and domain specificity of students' affect in the context of Latin instruction'. *British Journal of Educational Psychology* 76 pp.289-308. DOI: 10.1348/000709905X42860.

Goldstein, H., Browne, W. and Rasbash, J. (2002) 'Partitioning Variation in Multilevel Models'. *Understanding Statistics* 1(4) pp.223-231. DOI: 10.1207/S15328031US0104_02.

Google LLC. (2017) Android Studio, version 3.2.1. Available at: https://developer.android.com/studio.

Green Party. (2019) If Not Now, When? *Manifesto 2019.* London.

Grieco, L.A., Jowers, E.M. and Bartholomew, J.B. (2009) 'Physically active academic lessons and time on task: the moderating effect of body mass index'. *Medicine and Science in Sports and Exercise* 41(10) pp.1921-1926. DOI: 10.1249/MSS.0b013e3181a61495.

Grieco, L.A., Jowers, E.M., Errisuriz, V.L. and Bartholomew, J.B. (2016) 'Physically active vs. sedentary academic lessons: A dose response study for elementary student time on task'. *Preventive medicine* 89 pp.98-103. DOI: 10.1016/j.ypmed.2016.05.021.

Griffiths, L.J., Cortina-Borja, M., Sera, F., Pouliou, T., Geraci, M., Rich, C., Cole, T.J., Law, C., Joshi, H., Ness, A.R., Jebb, S.A. and Dezateux, C. (2013) 'How active are our children? Findings from the Millennium Cohort Study'. *BMJ Open* 3(8) pp.e002893. DOI: 10.1136/bmjopen-2013-002893.

Guiney, H. and Machado, L. (2013) 'Benefits of regular aerobic exercise for executive functioning in healthy populations'. *Psychonomic Bulletin and Review* 20(1) pp.73-86. DOI: 10.3758/s13423-012-0345-4.

Haapala, E.A., Vaisto, J., Lintu, N., Westgate, K., Ekelund, U., Poikkeus, A.M., Brage, S. and Lakka, T.A. (2016) 'Physical activity and sedentary time in relation to academic achievement in children'. *Journal of Science and Medicine in Sport* 20(6) pp.583-589. DOI: 10.1016/j.jsams.2016.11.003.

Hamaker, E.L., Nesselroade, J.R. and Molenaar, P.C.M. (2007) 'The integrated trait–state model'. *Journal of research in personality* 41(2) pp.295-315. DOI: 10.1016/j.jrp.2006.04.003.

Harris, J., Cale, L. and Musson, H. (2011) 'The effects of a professional development programme on primary school teachers' perceptions of physical education'. *Professional Development in Education* 37(2) pp.291-305. DOI: 10.1080/19415257.2010.531973.

Harris, J., Cale, L. and Musson, H. (2012) 'The predicament of primary physical education: a consequence of 'insufficient' ITT and 'ineffective' CPD?'. *Physical Education and Sport Pedagogy* 17(4) pp.367-381. DOI: 10.1080/17408989.2011.582489.

Harris, P. (2010) *Designing and reporting experiments in psychology,* Maidenhead: Open University Press.

Hayes, A.F. (2018) *Introduction to mediation, moderation, and conditional process analysis : a regression-based approach,* New York: The Guilford Press.

Hayes, A.F. and Preacher, K.J. (2014) 'Statistical mediation analysis with a multicategorical independent variable'. *British Journal of Mathematical and Statistical Psychology* 67(3) pp.451-470. DOI: 10.1111/bmsp.12028.

Heemskerk, C. (2015) 'Physical Education at Key Stage 2: Part of play deprivation or rebound? The relationship between classroom concentration and physical activity'. MSc Thesis, Univeristy of Oxford, Oxford.

Hillman, C.H., Castelli, D.M. and Buck, S.M. (2005) 'Aerobic Fitness and Neurocognitive Function in Healthy Preadolescent Children'. *Medicine & Science in Sports & Exercise* 37(11) pp.1967-1974. DOI: 10.1249/01.mss.0000176680.79702.ce.

Hillman, C.H., Erickson, K.I. and Kramer, A.F. (2008) 'Be smart, exercise your heart: exercise effects on brain and cognition.'. *Nature* 9 pp.58-65. DOI: 10.1038/nrn2298.

Hirvonen, R., Georgiou, G.K., Lerkkanen, M.K., Aunola, K. and Nurmi, J.E. (2010) 'Task-focused behaviour and literacy development: a reciprocal relationship'. *Journal of Research in Reading* 33(3) pp.302-319. DOI: 10.1111/j.1467-9817.2009.01415.x.

Hoetting, K., Schickert, N., Kaiser, J., Roder, B. and Schmidt-Kassow, M. (2016) 'The Effects of Acute Physical Exercise on Memory, Peripheral BDNF, and Cortisol in Young Adults'. *Neural Plasticity* 2016. DOI: 10.1155/2016/6860573.

Howie, E.K., Beets, M.W. and Pate, R.R. (2014) 'Acute classroom exercise breaks improve on-task behavior in 4th and 5th grade students: A dose–response'. *Mental Health and Physical Activity* 7(2) pp.65-71. DOI: 10.1016/j.mhpa.2014.05.002.

Hwang, J., Brothers, R.M., Castelli, D.M., Glowacki, E.M., Chen, Y.T., Salinas, M.M., Kim, J., Jung, Y. and Calvert, H.G. (2016) 'Acute high-intensity exercise-induced cognitive enhancement and brain-derived neurotrophic factor in young, healthy adults'. *Neuroscience Letters* 630 pp.247-253. DOI: 10.1016/j.neulet.2016.07.033.

IBM Corporation. (2016) IBM SPSS Statistics for Windows, version 24.0. Available at.

Inkinen, M., Lonka, K., Hakkarainen, K., Muukkonen, H., Litmanen, T. and Salmela-Aro, K. (2013) 'The Interface Between Core Affects and the Challenge– Skill Relationship'. *Journal of Happiness Studies* 15 pp.891-913. DOI: 10.1007/s10902-013-9455-6.

Iserbyt, P., Ward, P. and Li, W. (2017) 'Effects of improved content knowledge on pedagogical content knowledge and student performance in physical education'. *Physical Education and Sport Pedagogy* 22(1) pp.71-88. DOI: 10.1080/17408989.2015.1095868.

Jäger, K., Schmidt, M., Conzelmann, A. and Roebers, C.M. (2014) 'Cognitive and physiological effects of an acute physical activity intervention in elementary school children'. *Frontiers in Psychology* 5 pp.1473. DOI: 10.3389/fpsyg.2014.01473.

Jäger, K., Schmidt, M., Conzelmann, A. and Roebers, C.M. (2015) 'The effects of qualitatively different acute physical activity interventions in real-world settings on executive functions in preadolescent children'. *Mental Health and Physical Activity* 9 pp.1-9. DOI: 10.1016/j.mhpa.2015.05.002.

Janssen, M., Toussaint, H.M., van Mechelen, W. and Verhagen, E.A. (2014) 'Effects of acute bouts of physical activity on children's attention: a systematic review of the literature'. *SpringerPlus* 3 pp.410-419. DOI: 10.1186/2193-1801-3-410.

Jarrett, O.S., Maxwell, D.M., Dickerson, C., Hoge, P., Davies, G. and Yetley, A. (1998) 'Impact of Recess on Classroom Behavior: Group Effects and Individual Differences'. *The Journal of Educational Research* 92(2) pp.121-126. DOI: 10.1080/00220679809597584.

Jewitt, C. (2012) An introduction to using video for research - working paper. London: Institute of Education, National Centre for Research Methods.

Knaepen, K., Goekint, M., Heyman, E.M. and Meeusen, R. (2010) 'Neuroplasticity – Exercise-Induced Response of Peripheral Brain-Derived Neurotrophic Factor'. *Sports Medicine* 40(9) pp.765-801. DOI: 0112-1642/10/0009-0765.

Labour Party. (2019) It's time for real change. The Labour Party manifesto 2019. London.

Lambourne, K. and Tomporowski, P. (2010) 'The effect of exercise-induced arousal on cognitive task performance: a meta-regression analysis'. *Brain Research* 1341 pp.12-24. DOI: 10.1016/j.brainres.2010.03.091.

Leckie, G. and French, R. (2013) Three-Level Multilevel Models – MLwiN Practical. *LEMMA VLE Module 11*. (Accessed: Access Date).

Lees, C. and Hopkins, J. (2013) 'Effect of Aerobic Exercise on Cognition, Academic Achievement, and Psychosocial Function in Children: A Systematic Review of Randomized Control Trials'. *Preventing Chronic Disease* 10(10). DOI: 10.5888/pcd10.130010.

Lenhard, W. and Lenhard, A. (2016) *Calculation of Effect Sizes*. Available at: https://www.psychometrica.de/effect_size.html (Accessed: 15.2.2017).

Liberal Democrats. (2019) Stop Brexit. Build a brighter future. *Manifesto 2019.* 1.2 ed. London: Full Spectrum Print Media.

Long, B.C. and Van Stavel, R. (1995) 'Effects of exercise training on anxiety: A meta-analysis'. *Journal of Applied Sport Psychology* 7(2) pp.167-189. DOI: 10.1080/10413209508406963.

Lubans, D., Richards, J., Hillman, C., Faulkner, G., Beauchamp, M., Nilsson, M., Kelly, P., Smith, J., Raine, L. and Biddle, S. (2016) 'Physical Activity for Cognitive and Mental Health in Youth: A Systematic Review of Mechanisms'. *Pediatrics* 138(3) pp.1642-1656. DOI: 10.1542/peds.2016-1642.

Lubans, D.R., Lonsdale, C., Cohen, K., Eather, N., Beauchamp, M.R., Morgan, P.J., Sylvester, B.D. and Smith, J.J. (2017) 'Framework for the design and delivery of organized physical activity sessions for children and adolescents: rationale and description of the 'SAAFE' teaching principles'. *The international journal of behavioral nutrition and physical activity* 14(1) pp.24-34. DOI: 10.1186/s12966-017-0479-x.

Lundbye-Jensen, J., Skriver, K., Nielsen, J.B. and Roig, M. (2017) 'Acute Exercise Improves Motor Memory Consolidation in Preadolescent Children'. *Frontiers in Human Neuroscience* 11. DOI: 10.3389/fnhum.2017.00182.

Ma, J.K., Mare, L.L. and Gurd, B.J. (2014) 'Classroom-based high-intensity interval activity improves off-task behaviour in primary school students'. *Applied Physiology, Nutrition & Metabolism* 39(12) pp.1332-1337. DOI: dx.doi.org/10.1139/apnm-2014-0125.

Mabry, L. (2008) Case Study in Social Research. In: Alasuutari, P., Bickman, L. and Brannen, J. (eds) *The SAGE handbook of social research methods.* London: SAGE.

Macdonald, D., Abbott, R., Hunter, L., Hay, P. and McCuaig, L. (2014) 'Physical activity – academic achievement: student and teacher perspectives on the 'new' nexus'. *Physical Education and Sport Pedagogy* 19(4) pp.436-449. DOI: 10.1080/17408989.2013.769510.

Mackworth, J.F. (1965) 'The effect of amphetamine on the detectability of signals in a vigilance task'. *Canadian Journal of Psychology* 19(2) pp.104-110. DOI: 10.1037/h0082897.

Mackworth, N.H. (1948) 'The breakdown of vigilance during prolonged visual search'. *Quarterly Journal of Experimental Psychology* 1(1) pp.6-21. DOI: 10.1080/17470214808416738.

Mahar, M.T. (2019) 'Classroom-Based Physical Activity and On-Task Behavior'. *Translational Journal of the American College of Sports Medicine* 4(17) pp.148-154. DOI: 10.1249/tjx.0000000000000099.

Mahar, M.T., Murphy, S.K., Rowe, D.A., Golden, J., Shields, A.T. and Raedeke, T.D. (2006) 'Effects of a classroom-based program on physical activity and on-task behavior'. *Medicine and Science in Sports and Exercise* 38(12) pp.2086-2094. DOI: 10.1249/01.mss.0000235359.16685.a3.

Malmberg, L.-E., de Vos, M. and Saunders, K. (2016) Activity, rest and learning experiences (ARLE). John Fell Fund, 1.5.2016-31.7.2017.

Malmberg, L.-E. and Little, T.D. (2007) 'Profiles of ability, effort, and difficulty: Relationships with worldviews, motivation and adjustment'. *Learning and Instruction* 17(6) pp.739-754. DOI: https://doi.org/10.1016/j.learninstruc.2007.09.014.

Malmberg, L.-E., Pakarinen, E., Vasalampi, K. and Nurmi, J.-E. (2015) 'Students' school performance, task-focus, and situation-specific motivation'. *Learning and Instruction* 39 pp.158-167. DOI: 10.1016/j.learninstruc.2015.05.005.

Malmberg, L.-E., Walls, T.A., Martin, A.J., Little, T.D. and Lim, W.H.T. (2013a) 'Primary school students' learning experiences of, and self-beliefs about competence, effort, and difficulty: Random effects models'. *Learning and Individual Differences* 28 pp.54-65. DOI: https://doi.org/10.1016/j.lindif.2013.09.007.

Malmberg, L.-E., Woolgar, C. and Martin, A.J. (2013b) 'Quality of measurement of the learning experience questionnaire for personal digital assistants'. *International journal of quantitative research in education* 1(3) pp.275-296.

Martin, A.J. (2007) 'Examining a multidimensional model of student motivation and engagement using a construct validation approach'. *The British journal of educational psychology* (77) pp.413-440. DOI: 10.1348/000709906X118036.

Martin, A.J., Papworth, B., Ginns, P., Malmberg, L.-E., Collie, R.J. and Calvo, R.A. (2015) 'Real-time motivation and engagement during a month at school: Every moment of every day for every student matters'. *Learning and Individual Differences* 38(Supplement C) pp.26-35. DOI: https://doi.org/10.1016/j.lindif.2015.01.014.

Mavilidi, M.F., Drew, R., Morgan, P.J., Lubans, D.R., Schmidt, M. and Riley, N. (2019) 'Effects of different types of classroom physical activity breaks on children's on-task behaviour and academic achievement, and cognition'. *Acta paediatrica (Oslo, Norway: 1992)*. DOI: 10.1111/apa.14892.

Mavilidi, M.F., Ruiter, M., Schmidt, M., Okely, A.D., Loyens, S., Chandler, P. and Paas, F. (2018) 'A Narrative Review of School-Based Physical Activity for Enhancing Cognition and Learning: The Importance of Relevancy and Integration'. *Frontiers in Psychology* 9. DOI: 10.3389/fpsyg.2018.02079.

McHugh, M.L. (2012) 'Interrater reliability: the kappa statistic'. *Biochemia Medica* 22(3) pp.276-282. DOI: 10.11613/BM.2012.031.

McLellan, G., Arthur, R. and Buchan, D.S. (2018) 'Wear compliance, sedentary behaviour and activity in free-living children from hip-and wrist-mounted ActiGraph GT3X+ accelerometers'. *Journal of Sports Sciences* 36(21) pp.2424-2430. DOI: 10.1080/02640414.2018.1461322.

Meany-Daboul, M.G., Roscoe, E.M., Bourret, J.C. and Ahearn, W.H. (2007) 'A comparison of momentary time sampling and partial-interval recording for evaluating functional relations'. *Journal of Applied Behavior Analysis* 40(3) pp.501-514. DOI: 10.1901/jaba.2007.40-501.

Meinhardt, J. and Pekrun, R. (2003) 'Attentional resource allocation to emotional events: An ERP study'. *Cognition and Emotion* 17(3) pp.477-500. DOI: 10.1080/02699930244000039.

Merz, E.L. and Roesch, S.C. (2011) 'Modeling trait and state variation using multilevel factor analysis with PANAS daily diary data'. *Journal of research in personality* 45(1) pp.2. DOI: 10.1016/j.jrp.2010.11.003.

Migueles, J., Rowlands, A., Huber, F., Sabia, S. and Van Hees, V. (2019) 'GGIR: A Research Community-Driven Open Source R Package for Generating Physical Activity and Sleep Outcomes From Multi-Day Raw Accelerometer Data'. *Journal for the Measurement of Physical Behaviour* (2) pp.188-196. DOI: 10.1123/jmpb.2018-0063.

Miyake, A., Friedman, N.P., Emerson, M.J., Witzki, A.H., Howerter, A. and Wager, T.D. (2000) 'The unity and diversity of executive functions and their contributions to complex "Frontal Lobe" tasks: a latent variable analysis'. *Cognitive Psychology* 41(1) pp.49-100. DOI: 10.1006/cogp.1999.0734.

Morgan, P.J. and Hansen, V. (2008a) 'Classroom Teachers' Perceptions of the Impact of Barriers to Teaching Physical Education on the Quality of Physical Education Programs'. *Research Quarterly for Exercise and Sport* 79(4) pp.506-516. DOI: 10.1080/02701367.2008.10599517.

Morgan, P.J. and Hansen, V. (2008b) 'Physical education in primary schools: Classroom teachers' perceptions of benefits and outcomes'. *Health Education Journal* 67(3) pp.196-207. DOI: 10.1177/0017896908094637.

Muthen, B. and Muthen, L. (2017) Mplus version 8. Available at: http://www.statmodel.com/orderonline/.

Muthén, B., O. (1991) 'Multilevel Factor Analysis of Class and Student Achievement Components'. *Journal of Educational Measurement* 28(4) pp.338-354. DOI: 10.1111/j.1745-3984.1991.tb00363.x.

Muthén, B.O. (1994) 'Multilevel Covariance Structure Analysis'. *Sociological Methods & Research* 22(3) pp.376-398. DOI: 10.1177/0049124194022003006.

Newton, D.P. (2013) 'Moods, emotions and creative thinking: A framework for teaching'. *Thinking Skills and Creativity* 8(Supplement C) pp.34-44. DOI: https://doi.org/10.1016/j.tsc.2012.05.006.

Nieuwenhuis, S., Aston-Jones, G. and Cohen, J.D. (2005) 'Decision making, the P3, and the locus coeruleus-norepinephrine system'. *Psychological Bulletin* 131(4) pp.510-532. DOI: 10.1037/0033-2909.131.4.510.

Ofsted. (2014) The PE and sport premium for primary schools. Manchester: Ofsted.

Parfitt, G. and Eston, R.G. (2005) 'The relationship between children's habitual activity level and psychological well-being'. *Acta Pædiatrica* 94(12) pp.1791-1797. DOI: 10.1111/j.1651-2227.2005.tb01855.x.

Pearson. (2013) BOSS User's Guide. Bloomington: Pearson.

Pearson. (2019) *Behavior Assessment System for Children - Third Edition*. Available at: https://www.pearsonassessments.com/store/usassessments/en/Store/Professional-Assessments/Behavior/Comprehensive/Behavior-Assessment-System-for-Children-%7C-Third-Edition-/p/100001402.html?tab=product-details (Accessed: 29.10.2019).

Pekrun, R. (2006) 'The Control-Value Theory of Achievement Emotions: Assumptions, Corollaries, and Implications for Educational Research and Practice'. *Educational Psychology Review* 18(4) pp.315-341. DOI: 10.1007/s10648-006-9029-9.

Pekrun, R., Goetz, T., Titz, W. and Perry, R.P. (2002) 'Academic emotions in students' self-regulated learning and achievement: A program of qualitative and quantitative research'. *Educational Psychologist* 37(2) pp.91-105. DOI: 10.1207/S15326985EP3702_4.

Pellegrini, A.D. and Davis, P. (1993) 'Relations between children's playground and classroom behaviour'. *British Journal of Educational Psychology* 63(1) pp.88-95. DOI: 10.1111/j.2044-8279.1993.tb01043.x.

Pellegrini, A.D. and Holmes, R.M. (2006) The Role of Recess in Primary School. In: Singer, D.G., Golinkoff, R.M. and Hirsh-Pasek, K. (eds) *Play = Learning: How Play Motivates and Enhances Children's Cognitive and Social-Emotional Growth.* Oxford: Oxford Scholarship Online, pp.36-56.

Pellegrini, A.D., Huberty, P.D. and Jones, I. (1995) 'The Effects of Recess Timing on Children's Playground and Classroom Behaviors'. *American Educational Research Journal* 32(4) pp.845-864. DOI: 10.3102/00028312032004845.

Pellegrini, A.D. and Smith, P.K. (1993) 'School Recess: Implications for Education and Development'. *Review of Educational Research* 63(1) pp.51-67. DOI: 10.3102/00346543063001051.

Pesce, C., Crova, C., Cereatti, L., Casella, R. and Bellucci, M. (2009) 'Physical activity and mental performance in preadolescents: Effects of acute exercise on free-recall memory'. *Mental Health and Physical Activity* 2(1) pp.16-22. DOI: https://doi.org/10.1016/j.mhpa.2009.02.001.

Phillips, L.R., Parfitt, G. and Rowlands, A.V. (2013) 'Calibration of the GENEA accelerometer for assessment of physical activity intensity in children'. *Journal of Science and Medicine in Sport* 16(2) pp.124-128. DOI: 10.1016/j.jsams.2012.05.013.

Pimentel, J., Tojal, R. and Morgado, J. (2015) 'Epilepsy and physical exercise'. *Seizure* 25 pp.87-94. DOI: https://doi.org/10.1016/j.seizure.2014.09.015.

Poh, M.Z., Swenson, N.C. and Picard, R.W. (2010) 'A wearable sensor for unobtrusive, long-term assessment of electrodermal activity'. *IEEE Transactions on Biomedical Engineering* 57(5) pp.1243-1252. DOI: 10.1109/TBME.2009.2038487.

Pontifex, M.B., Raine, L.B., Johnson, C.R., Chaddock, L., Voss, M.W., Cohen, N.J., Kramer, A.F. and Hillman, C.H. (2011) 'Cardiorespiratory fitness and the flexible modulation of cognitive control in preadolescent children'. *Journal of Cognitive Neuroscience* 23(6) pp.1332-1345. DOI: 10.1162/jocn.2010.21528.

Posner, M.I., Rueda, M.R. and Kanske, P. (2007) Probing the Mechanisms of Attention. In: Cacioppo, J.T., Tassinary, L.G. and Berntson, G.G. (eds) *Handbook of Psychophysiology.* 3rd ed. Cambridge: Cambridge University Press, pp.410-432.

Pöysä, S., Vasalampi, K., Muotka, J., Lerkkanen, M.-K., Poikkeus, A.-M. and Nurmi, J.-E. (2017) 'Variation in situation-specific engagement among lower secondary school students'. *Learning and Instruction* 53(Supplement C) pp.64-73. DOI: https://doi.org/10.1016/j.learninstruc.2017.07.007.

Preacher, K.J., Rucker, D.D. and Hayes, A.F. (2007) 'Addressing Moderated Mediation Hypotheses: Theory, Methods, and Prescriptions'. *Multivariate Behavioral Research* 42(1) pp.185-227. DOI: 10.1080/00273170701341316.

Rachele, J., Cuddihy, T., Washington, T. and McPhail, S. (2016) 'School-Based Youth Physical Activity Promotion: Thoughts and Beliefs of Pre-Service Physical Education Teachers'. *Australian Journal of Teacher Education* 41(5) pp.52-64. DOI: 10.14221/ajte.2016v41n5.4.

Rainer, P., Cropley, B., Jarvis, S. and Griffiths, R. (2012) 'From policy to practice: the challenges of providing high quality physical education and school sport faced by head teachers within primary schools'. *Physical Education and Sport Pedagogy* 17(4) pp.429-446. DOI: 10.1080/17408989.2011.603125.

Rane, S. (2017) EmojiOne. Available at: https://www.emojione.com/ (Downloaded: 16.11.2017).

Robbins, T.W. (2002) 'The 5-choice serial reaction time task: behavioural pharmacology and functional neurochemistry'. *Psychopharmacology (Berl)* 163(3-4) pp.362-380. DOI: 10.1007/s00213-002-1154-7.

Rock, M.L. (2005) 'Use of Strategic Self-Monitoring to Enhance Academic Engagement, Productivity, and Accuracy of Students With and Without Exceptionalities'. *Journal of Positive Behavior Interventions* 7(1) pp.3-17. DOI: 10.1177/10983007050070010201.

Roig, M., Nordbrandt, S., Geertsen, S.S. and Nielsen, J.B. (2013) 'The effects of cardiovascular exercise on human memory: a review with meta-analysis'. *Neuroscience and Biobehavioral Reviews* 37(8) pp.1645-1666. DOI: 10.1016/j.neubiorev.2013.06.012.

Romero, I. (2006) The human nervous system. In: Bearman, G., Graham, R., Riley, G. and Wardell, P. (eds) *From cells to consciousness.* Milton Keynes: The Open University, pp.85-138.

Roorda, D.L., Koomen, H.M.Y., Spilt, J.L. and Oort, F.J. (2011) 'The Influence of Affective Teacher–Student Relationships on Students' School Engagement and Achievement'. *Review of Educational Research* 81(4) pp.493-529. DOI: 10.3102/0034654311421793.

Rosenthal, R. and Jacobson, L. (1968) 'Pygmalion in the classroom'. *The Urban Review* 3(1) pp.16-20. DOI: 10.1007/BF02322211.

Rothwell, P.M. (2005) 'External validity of randomised controlled trials: "To whom do the results of this trial apply?"'. *The Lancet* 365(9453) pp.82-93. DOI: http://dx.doi.org/10.1016/S0140-6736(04)17670-8.

Rowland, T.W. (2008) Cardiovascular function. In: Armstrong, N. and Van Mechelen, W. (eds) *Paediatric Exercise Science and Medicine.* 2nd ed. Oxford: Oxford University Press.

Royal College of Peadiatrics and Child Health. (2013) Boys and Girls BMI Chart. Royal College of Paediatrics and Child Health and Department of Health.

Rusch, H. and Irrgang, W. (no date) Muenchner Fitnesstest. Muenchen: Technischen Universitaet Muenchen.

Rush, J. and Hofer, S.M. (2014) 'Differences in within- and between-person factor structure of positive and negative affect: analysis of two intensive measurement studies using multilevel structural equation modeling'. *Psychological assessment* 26(2) pp.462. DOI: 10.1037/a0035666.

Sabia, S., van Hees, V.T., Shipley, M.J., Trenell, M.I., Hagger-Johnson, G., Elbaz, A., Kivimaki, M. and Singh-Manoux, A. (2014) 'Association Between Questionnaire- and Accelerometer-Assessed Physical Activity: The Role of Sociodemographic Factors'. *American Journal of Epidemiology* 179(6) pp.781-790. DOI: 10.1093/aje/kwt330.

Salis, A.S. (2013) 'Proactive and Reactive Effects of Vigorous Exercise on Learning and Vocabulary Comprehension'. *Perceptual and Motor Skills* 116(3) pp.918-928. DOI: 10.2466/29.22.PMS.116.3.918-928.

Sanders, T. (2019) iPLAY GENEActiv processing data - base script. Personal communication 27.3.

Schaefer, C.A., Nigg, C.R., Hill, J.O., Brink, L.A. and Browning, R.C. (2014) 'Establishing and evaluating wrist cutpoints for the GENEActiv accelerometer in youth'. *Medicine & Science in Sports & Exercise* 46(4) pp.826-833. DOI: 10.1249/MSS.0000000000000150.

Schmidt-Kassow, M., Deusser, M., Thiel, C., Otterbein, S., Montag, C., Reuter, M., Banzer, W. and Kaiser, J. (2013) 'Physical exercise during encoding improves vocabulary learning in young female adults: a neuroendocrinological study'. *PLoS One* 8(5). DOI: 10.1371/journal.pone.0064172.

Schmidt-Kassow, M., Zink, N., Mock, J., Thiel, C., Vogt, L., Abel, C. and Kaiser, J. (2014) 'Treadmill walking during vocabulary encoding improves verbal long-term memory'. *Behav Brain Funct* 10. DOI: 10.1186/1744-9081-10-24.

Schmidt, M., Benzing, V., Bonadimann, P. and Martin-Niedecken, A.L. (2019) 'Cognitively engaging physical activity and cognive performance: is there a dose-response relationship?', *Healthy and Active Children.* Verona, 12.09.2019. I-MDRC & CIAPSE.

Schmidt, M., Benzing, V. and Kamer, M. (2016) 'Classroom-based physical activity breaks and children's attention: cognitive engagement works'. *Frontiers in Psychology* 7. DOI: 10.3389/fpsyg.2016.01474.

Schmidt, M., Egger, F. and Conzelmann, A. (2015a) 'Delayed Positive Effects of an Acute Bout of Coordinative Exercise on Children's Attention'. *Perceptual and Motor Skills* 121(2) pp.431-446. DOI: 10.2466/22.06.PMS.121c22x1.

Schmidt, M., Jager, K., Egger, F., Roebers, C.M. and Conzelmann, A. (2015b) 'Cognitively Engaging Chronic Physical Activity, but not Aerobic Exercise, Affects Executive Functions in Primary School Children: A Group-Randomized Controlled Trial'. *Journal of Sport and Exercise Psychology* 37(6) pp.575-591. DOI: 10.1123/jsep.2015-0069.

Schmidt, R.A. and Wrisberg, C.A. (2004) *Motor Learning and Performance,* Leeds: Human Kinetics.

Scrimin, S. and Mason, L. (2015) 'Does mood influence text processing and comprehension? Evidence from an eye-movement study'. *British Journal of Educational Psychology* 85(3) pp.387-406. DOI: 10.1111/bjep.12080.

Sedgwick, P. (2013) 'Convenience sampling'. *BMJ: British Medical Journal* 347. DOI: 10.1136/bmj.f6304.

Singh, A.S., Saliasi, E., van Den Berg, V., Uijtdewilligen, L., de Groot, R.H.M., Jolles, J., Andersen, L.B., Bailey, R., Chang, Y.-K., Diamond, A., Ericsson, I., Etnier, J.L., Fedewa, A.L., Hillman, C.H., McMorris, T., Pesce, C., Pühse, U., Tomporowski, P.D. and Chinapaw, M.J.M. (2019) 'Effects of

physical activity interventions on cognitive and academic performance in children and adolescents: a novel combination of a systematic review and recommendations from an expert panel'. *British Journal of Sports Medicine* 53(10) pp.640. DOI: 10.1136/bjsports-2017-098136.

Skinner, E., Furrer, C., Marchand, G. and Kindermann, T. (2008) 'Engagement and Disaffection in the Classroom: Part of a Larger Motivational Dynamic?'. *Journal of Educational Psychology* 100(4) pp.765-781. DOI: 10.1037/a0012840.

Smith, P.K. and Hagan, T. (1980) 'Effects of Deprivation on Exercise Play in Nursery School Children'. *Animal Behaviour* 28(3) pp.922-928. DOI: 10.1016/S0003-3472(80)80154-0.

Steinlin, M. (2007) 'The cerebellum in cognitive processes: supporting studies in children'. *Cerebellum* 6(3) pp.237-241. DOI: 10.1080/14734220701344507.

Strauss, R.S., Rodzilsky, D., Burack, G. and Colin, M. (2001) 'Psychosocial correlates of physical activity in healthy children'. *Archives of Pediatrics & Adolescent Medicine* 155(8) pp.897-902. DOI: 10.1001/archpedi.155.8.897.

Sulkunen, P. (2012) Social Research and Social Practice in Post-Positivist Society. In: Alasuutari, P., Bickman, L. and Brannen, J. (eds) *The SAGE Handbook of Social Research Methods.* London: SAGE Publications Ltd, pp.68-80.

Szabo-Reed, A.N., Willis, E.A., Lee, J., Hillman, C.H., Washburn, R.A. and Donnelly, J.E. (2017) 'Impact of Three Years of Classroom Physical Activity Bouts on Time-on-Task Behavior'. *Medicine & Science in Sports & Exercise* 49(11) pp.2343-2350. DOI: 10.1249/MSS.0000000000001346.

Tabachnick, B.G. and Fidell, L.S. (2013) *Using multivariate statistics,* Boston/London: Pearson.

The Daily Mile Foundation. (2019) *The Daily Mile, children fit for life*. Available at: https://thedailymile.co.uk/ (Accessed: 6.11.2019).

Tomporowski, P.D. (2003) 'Effects of acute bouts of exercise on cognition'. *Acta Psychologica* 112(3) pp.297-324. DOI: 10.1016/S0001-6918(02)00134-8.

Tsai, C.L., Chen, F.C., Pan, C.Y., Wang, C.H., Huang, T.H. and Chen, T.C. (2014) 'Impact of acute aerobic exercise and cardiorespiratory fitness on visuospatial attention performance and serum BDNF levels'. *Psychoneuroendocrinology* 41 pp.121-131. DOI: 10.1016/j.psyneuen.2013.12.014.

Tsangaridou, N. (2017) 'Early childhood teachers' views about teaching physical education: challenges and recommendations'. *Physical Education and Sport Pedagogy* 22(3) pp.283-300. DOI: 10.1080/17408989.2016.1192593.

Valentini, N.C., Pierosan, L., Rudisill, M.E. and Hastie, P.A. (2017) 'Mastery and exercise play interventions: motor skill development and verbal recall of children with and without disabilities'. *Physical Education and Sport Pedagogy* 22(4) pp.349-363. DOI: 10.1080/17408989.2016.1241223.

Van der Niet, A.G., Hartman, E., Smith, J. and Visscher, C. (2014) 'Modeling relationships between physical fitness, executive functioning, and academic achievement in primary school children'. *Psychology of Sport and Exercise* 15(4) pp.319-325. DOI: 10.1016/j.psychsport.2014.02.010.

Van Hees, V.T., Fang, Z., Langford, J., Assah, F., Mohammad, A., Da Silva, I.C.M., Trenell, M.I., White, T., Wareham, N.J. and Brage, S. (2014) 'Autocalibration of accelerometer data for free-living physical activity assessment using local gravity and temperature: an evaluation on four continents'. *Journal of Applied Physiology* 117(7) pp.738-744. DOI: 10.1152/japplphysiol.00421.2014.

Vasalampi, K., Muotka, J., Pöysä, S., Lerkkanen, M.-K., Poikkeus, A.-M. and Nurmi, J.-E. (2016) 'Assessment of students' situation-specific classroom engagement by an InSitu Instrument'. *Learning and Individual Differences* 52(Supplement C) pp.46-52. DOI: https://doi.org/10.1016/j.lindif.2016.10.009.

Verburgh, L., Königs, M., Scherder, E.J.A. and Oosterlaan, J. (2014) 'Physical exercise and executive functions in preadolescent children, adolescents and young adults: a meta-analysis'. *British Journal of Sports Medicine* 48(12) pp.973. DOI: 10.1136/bjsports-2012-091441.

Voelcker-Rehage, C. and Niemann, C. (2013) 'Structural and functional brreain changes related to different types of physical activity across the life span'. *Neurosci. Biobehav. Rev.* 37(9) pp.2268-2295. DOI: 10.1016/j.neubiorev.2013.01.028.

Wake Up Shake Up Productions. (2014) *Wake Up! Shake Up!* Available at: http://www.wakeupshakeup.com/ (Accessed: 6/11/2019).

Watson, A., Timperio, A., Brown, H., Best, K. and Hesketh, K.D. (2017) 'Effect of classroom-based physical activity interventions on academic and physical activity outcomes: a systematic review and meta-analysis'. *International Journal of Behavioral Nutrition and Physical Activity* 14(1) pp.114. DOI: 10.1186/s12966-017-0569-9.

Watson, A., Timperio, A., Brown, H., Hinkley, T. and Hesketh, K.D. (2019) 'Associations between organised sport participation and classroom behaviour outcomes among primary school-aged children'. *PLoS One* 14(1). DOI: 10.1371/journal.pone.0209354.

Watson, D., Clark, L. A., Tellegen, A. (1988) 'Development and Validation of Brief Measures of Positive and Negative Affect: The PANAS Scales'. *Journal of Personality and Social Psychology* 54(6) pp.1063-1070.

Weinberg, L., Hasni, A., Shinohara, M. and Duarte, A. (2014) 'A single bout of resistance exercise can enhance episodic memory performance'. *Acta Psychologica (Amsterdam)* 153 pp.13-19. DOI: 10.1016/j.actpsy.2014.06.011.

Westfall, D.R., Kao, S.C., Scudder, M.R., Pontifex, M.B. and Hillman, C.H. (2017) 'The association between aerobic fitness and congruency sequence effects in preadolescent children'. *Brain and Cognition* 113 pp.85-92. DOI: 10.1016/j.bandc.2016.12.005.

Whitt-Glover, M.C., Ham, S.A. and Yancey, A. (2011) 'Instant Recess (R): A Practical Tool for Increasing Physical Activity During the School Day'. *Progress in Community Health Partnerships: Research, Education, and Action* 5(3) pp.289-297.

Wilson, A., Olds, T., Lushington, K., Petkov, J. and Dollman, J. (2016) 'The impact of 10-minute activity breaks outside the classroom on male students' on-taskbehaviour and sustained attention: a randomised crossover design.'. *Acta Pædiatrica* 105(4) pp.E181-E188. DOI: 10.1111/apa.13323.

Winter, B., Breitenstein, C., Mooren, F.C., Voelker, K., Fobker, M., Lechtermann, A., Krueger, K., Fromme, A., Korsukewitz, C., Floel, A. and Knecht, S. (2007) 'High impact running improves learning'. *Neurobiology of Learning and Memory* 87(4) pp.597-609. DOI: 10.1016/j.nlm.2006.11.003.

Yerkes, R.M. and Dodson, J.D. (1908) 'The Relations of Strength of Stimulus to Rapidity of Habit-Formation'. *Journal of Comparative Neurology and Psychology* 18(5) pp.459-482. DOI: 10.1002/cne.920180503.

Yin, R.K. (2013) *Case study research: design and methods,* Los Angeles: SAGE.

Yu, C.C.W., Chan, S., Cheng, F., Sung, R.Y.T. and Hau, K.T. (2006) 'Are physical activity and academic performance compatible? Academic achievement, conduct, physical activity and self-esteem of Hong Kong Chinese primary school children'. *Educational Studies* 32(4) pp.331-341. DOI: 10.1080/03055690600850016.

Zakszeski, B.N., Hojnoski, R.L. and Wood, B.K. (2017) 'Considerations for Time Sampling Interval Durations in the Measurement of Young Children's Classroom Engagement'. *Topics In Early Childhood Special Education* 37(1) pp.42-53. DOI: 10.1177/0271121416659054.