School-based physical activity interventions, physical fitness and mental health among adolescents

Effects, associations and lessons learned from the School in Motion study

by

Andreas Åvitsland

Thesis submitted in fulfilment of
the requirements for the degree of
PHILOSOPHIAE DOCTOR
(PhD)

Faculty of Arts and Education
Department of Education and Sports science
2020

University of Stavanger
NO-4036 Stavanger
NORWAY
www.uis.no

ISBN: 978-82-7644-969-3
ISSN: 1890-1387
PhD: Thesis UiS No. 561

Acknowledgements

It has been three years since I was offered the opportunity to go to Stavanger and work on some of the most interesting research topics that I can think of. Today, although I have never been more tired of anything, I look back with gratitude for the chance that I was given.

Much gratitude and respect go to my always available supervisors and co-authors. They offered their guidance whenever it was needed.

Three years could not have gone by so successfully without the support from family and friends. You have provided me with welcome distractions during periods of monotonous focus.

Most importantly, the biggest thanks go to my wife (and my dog), who gave me the strength to persevere almost every day for three years. Thank you.

Andreas Åvitsland

Summary

This thesis is part of the School in Motion study. School in Motion was initiated by the Norwegian government because of the political demand for more knowledge regarding possible approaches to increase adolescent physical activity. The reason for the demand was a report depicting 15-year-olds as among the most sedentary populations in Norway. This is concerning because physical activity can have a preventive function against non-communicable diseases, psychological difficulties and mental disorders. Norwegian schools are given extended responsibilities for their students' physical and mental health. Physical activity can contribute in this purpose, therefore, finding feasible approaches to increase physical activity in school is an important research topic.

The School in Motion study was a cluster randomized controlled trial involving 29 lower secondary schools in Norway. The participants were 2084 9th graders (14-15 years old) from schools located in western, south western, southern, and eastern parts of the country. The study aimed to implement two separate intervention models for increased physical activity and physical education in school, in order to evaluate their effects on physical activity, physical health, mental health, learning and learning environment. Intervention model 1 included weekly, 30 minutes of physical activity, 30 minutes of physically active academic lessons, and one additional physical education lesson. Intervention model 2 included two weekly physical activity lessons in which students were encouraged to pursue their own activity interests, in groups they formed themselves.

The aims of this thesis were to elucidate different aspects of the relationship between school-based physical activity interventions, physical fitness and mental health among adolescents, and the feasibility of school-based PA interventions in lower secondary school. The data

that are presented in this thesis are the basis of four separate papers. The first paper analyzed baseline data to examine the cross-sectional association between physical fitness and mental health. The results showed a small, inverse association between cardiorespiratory fitness and mental health, while no association was found between mental health and body composition or muscular strength.

The second paper presents data from a qualitative study that aimed to elucidate how intervention model 2 was implemented, and what influenced the implementation. The results showed large differences in how the intervention was implemented. Specifically, two out of four investigated schools implemented the intervention adequately, one school implemented the intervention partially adequate and the last school failed at several implementation aspects. Four main factors were responsible for influencing implementation: frame factors, intervention-, participant-, and provider characteristics. The results elucidate the challenge of implementing a complex physical activity intervention in schools, i. e., complex contexts, and expecting the implementation process to be similar for the schools.

The third paper examined the effect the interventions had on mental health, by using a cluster randomized controlled trial design. The results showed no intervention effects on the overall population. However, subgroup analyses indicated that intervention model 1 was beneficial for those with high levels of psychological difficulties at baseline, and both intervention models were beneficial for the immigrant subgroup. The results also indicated that intervention model 2 had a negative effect on a subscale of psychological difficulties, peer relationships, within the subgroups non-immigrant girls, and those with "borderline" psychological difficulties at baseline.

The fourth paper presents associations between the one-year change in physical fitness components and mental health status. Subgroup analyses showed an inverse association between change in cardiorespiratory

fitness and mental health status among boys; an inverse association between change in muscular strength and mental health status among immigrants; and an association between change in cardiorespiratory fitness and mental health among girls with high socioeconomic status. The results indicate that the associations between physical activity, physical fitness and mental health among adolescents depend on different mechanisms that are specific to certain subgroups.

The findings of this thesis add nuance to a research field characterized by relatively few studies, small sample sizes, cross-sectional designs and little qualitative knowledge. In summary, this thesis suggests that improving cardiorespiratory fitness can be beneficial for mental health promotion among adolescent boys; and that improving muscular strength can be beneficial for mental health promotion among adolescent immigrants. Moreover, school-based PA programs can be useful for mental health promotion among adolescents with high levels of psychological difficulties and among adolescent immigrants. Regarding the implementation of physical activity programs in schools, this is feasible. However, a successful implementation depends to a large degree on schools' facilities and frame factors, thus, the physical activity programs should have flexible designs to secure an optimal fit in each individual school.

The new Core curriculum in Norway for primary and secondary education specifies that physical and mental health is to be an important interdisciplinary topic in school. The knowledge generated through this thesis can contribute to the development of this interdisciplinary topic, and to the ongoing discussion about the role of physical activity and mental health in school.

List of papers

Paper 1: Åvitsland, A., Leibinger, E., Haugen, T., Lerum, Ø., Solberg, R. B., Kolle, E., & Dyrstad, S. M. (2020) The association between mental health and physical fitness in Norwegian adolescents. *BMC Public Health, 20*. doi: https://doi.org/10.1186/s12889-020-08936-7

Paper 2: Åvitsland, A., Ohna, S. E., Dyrstad, S. M., Tjomsland, H. E., Lerum, Ø. & Leibinger, E. (2020) The process evaluation of a school-based physical activity intervention: Influencing factors and potential consequences of implementation. *Health Education, 120*(2). doi: https://doi.org/10.1108/HE-01-2020-0004

Paper 3: Åvitsland, A., Leibinger, E., Resaland, G. K., Solberg, R. B., Kolle, E. & Dyrstad, S. M. (2020) Effects of school-based physical activity interventions on mental health in adolescents: The School in Motion cluster randomized controlled trial. *Mental Health and Physical Activity, in press*. doi online journal pre-proof: https://doi.org/10.1016/j.mhpa.2020.100348

Paper 4: Åvitsland, A., Leibinger, E., Kolle, E., Haugen, T. & Dyrstad, S. M. (2020) An investigation of the prospective association between physical fitness and mental health in Norwegian adolescents. *Manuscript submitted for publication*

Abbreviations

ADHD: Attention deficit hyperactivity disorder

BMI: Body mass index

CRF: Cardiorespiratory fitness

DWBH: Don’t worry, be happy (intervention model 2)

MET: Metabolic equivalent

PA: Physical activity

PAAL: Physically active academic lessons

PE: Physical education

PME: Positive Movement Experiences

PYD: Positive Youth Development

RCT: Randomized controlled trial

RDS: Relational Developmental Systems

SES: Socioeconomic status

WHO: The World Health Organization

Table of Contents

Acknowledgements iv

Summary v

List of papers viii

Abbreviations ix

1 Definition of terms 1

1.1 Physical activity 1

1.2 Physical fitness 2

1.3 Mental health and mental disorders 2

1.4 Adolescence 3

2 Introduction 4

2.1 A brief history of physical activity and health in schools 5

2.2 The School in Motion study - background 7

2.3 Contribution and aims 8

3 Theoretical framework and empirical background 11

3.1 Mental health 11

3.2 Physical activity and fitness 12

3.3 Physical activity, physical fitness and mental health 13

3.3.1 Explanatory mechanisms 14

3.4 Evaluating complexity 18

3.5 Previous interventions 22

4 Research questions 25

5 Methods 26

5.1 Quantitative study designs 26

5.1.1 Participants 28

5.1.2 Interventions 28

5.1.3 Outcome measures 36

5.1.4 Statistical analyses (Papers 1, 3 and 4) 39

5.2 Qualitative study design 44

5.2.1 Participants 44

5.2.2 Data collection 45

5.2.3 Qualitative analysis (Paper 2) 46
5.3 Ethical considerations 47
5.3.1 School in Motion 47
5.3.2 The process evaluation 50
6 Results 53
6.1 Paper 1 53
6.2 Paper 2 54
6.3 Paper 3 55
6.4 Paper 4 57
7 Discussion 59
7.1 Research question 1 59
7.1.1 Cardiorespiratory fitness 60
7.1.2 Muscular strength 65
7.1.3 Body composition 67
7.2 Research question 2 68
7.3 Research question 3 71
7.3.1 Abnormal psychological difficulties at baseline 73
7.3.2 Immigrants 74
7.3.3 Detrimental effects 76
7.4 Methodological considerations 77
7.4.1 Quantitative studies 77
7.4.2 Qualitative study 81
7.5 Implications 81
7.6 Recommendations for future research 84
8 Conclusion 86
9 References 87
Paper 1 120
Paper 2 130
Paper 3 149
Paper 4 178
Appendices 203

List of Figures

Figure 1. Participant flow, missing data and participant distribution for each quantitative research paper.

Figure 2. School in Motion project timeline.

Appendices

Appendix 1: Approval for the School in Motion project from NSD (Norwegian Centre for Research Data)

Appendix 2: The Strengths and Difficulties Questionnaire

Appendix 3: The interview guide

Appendix 4: The activity contract for “Don’t worry, be happy”

Appendix 5: Supplementary table 1a (results from Paper 4)

1 Definition of terms

1.1 Physical activity

Physical activity (PA) can be defined as "any bodily movement produced by skeletal muscles that results in energy expenditure" (Caspersen, Powell, & Christenson, 1985). What determines the amount of energy expenditure is the intensity of the activity, which often is measured by metabolic equivalents (MET; Ainsworth et al., 2000). We define one MET as "the resting metabolic rate, that is, the amount of oxygen consumed at rest, sitting quietly in a chair" (Jetté, Sidney, & Blümchen, 1990) and one MET is equal to 3.5 ml oxygen per kg body weight per minute. METs are used to categorize physical activity intensity levels, and Pate et al. (1995) proposed the categories light intensity PA (<3.0 METs), moderate intensity PA (3.0-6.0 METs), and vigorous PA (>6.0 METs). To illustrate examples of PA intensities, Ainsworth et al. (2000) published a compendium outlining several different activity types and the corresponding METs, for instance vacuuming (3.5 METs), bicycling (8.0 METs) and standing (2.0 METs). We still use the three PA intensity categories, mostly for providing guidelines and recommendations regarding the population's PA levels. The World Health Organization (WHO) provide updated guidelines in their "Global recommendations on physical activity for health" (2010):

Adults aged 18–64 years should do at least 150 minutes of moderate-intensity aerobic physical activity throughout the week, or do at least 75 minutes of vigorous-intensity aerobic physical activity throughout the week, or an equivalent combination of moderate- and vigorous-intensity activity. (p. 8)

WHOs recommendations for children and adolescents (5-17 years old) are: "Children and young people aged 5–17 years old should accumulate at least 60 minutes of moderate to vigorous-intensity physical activity daily" (2010, p. 7).

1.2 Physical fitness

Physical fitness is a set of attributes that are present or can be achieved, based on type, duration and intensity of PA (Blair, Cheng, & Holder, 2001). Caspersen et al. (1985) divided the concept physical fitness into health-related fitness and skill-related fitness. Under skill-related fitness are the specific qualities agility, balance, coordination, speed, power and reaction time. Under health-related fitness are the aspects cardiorespiratory fitness, muscular endurance, muscular strength, body composition and flexibility. Going forward, this thesis will focus on cardiorespiratory fitness, muscular strength and body composition, and use the definitions from Caspersen et al. (1985): Cardiorespiratory fitness (CRF) is the ability of the circulatory and respiratory system to supply working muscles with oxygen during sustained physical activity. Muscular strength is the ability of the muscles to produce external force. Body composition is the relative amount of muscle, fat, bone and other parts of the body.

1.3 Mental health and mental disorders

The World Health Organization defines mental health as "a state of well-being in which every individual realizes his or her own potential, can cope with the normal stresses of life, can work productively and fruitfully, and is able to make a contribution to her or his community" (2014). Opposite of mental health, we find mental disorders, which is defined in the Diagnostic and statistical manual of mental disorders (DSM-5) by the American Psychiatric Association (2013):

A mental disorder is a syndrome characterized by clinically significant disturbance in an individual's cognition, emotion regulation, or behavior that reflects a dysfunction in the psychological, biological, or developmental processes underlying mental functioning. Mental disorders are usually associated with significant distress or disability in social, occupational, or other important activities. (p. 20).

1.4 Adolescence

Adolescence is generally described as the period that starts with puberty and lasts until the individual has established social independence (Steinberg, 2014). The adolescent age span is debated and consists of developmental sub-stages, which are also debated; however, the most commonly used age span for this period is from 10-18 years (Curtis, 2015). Adolescent development is characterized by major biological, social, physical, cognitive and psychological changes, which influences behavior and can influence how adolescents respond to interventions (Christie & Viner, 2005). Scholars have advocated to "de-dramatize" adolescence by arguing that although the period is characterized by problematic behavior, emotional difficulties and parental conflicts, the period passes without turmoil for most individuals (Dornbusch, Petersen, & Hetherington, 1991). However, since the research field of adolescent development emerged more than a century ago, the research has been directed towards portraying healthy adolescent development as "avoidance of problems" rather than "growth of competencies" (Steinberg & Morris, 2001, p. 85).

This thesis adheres to the tenets of contemporary research on adolescent development, which, according to Steinberg and Lerner (2004), focus on developing knowledge useful for practitioners and policy makers, for the purpose of facilitating positive development that will benefit society in general. "Above all, the study of adolescent development at its best both informs and is informed by the concerns of communities, of practitioners, and of policy makers" (Steinberg & Lerner, 2004, p. 52).

2 Introduction

There are several outlined objectives for the education of Norwegian children and adolescents ("The Education Act," 1998b § 1-1). In short, students shall acquire knowledge, skills, attitudes and competencies that are important for promoting “democracy, equality and scientific thinking”, as well as mastering their lives so that they can work and otherwise contribute to society. Thus, the course of primary and secondary education is primarily characterized by learning and development. However, ("The Education Act," 1998a, § 9-A-1) also states that students’ mental health is among the responsibilities of the education: “All pupils are entitled to a good physical and psychosocial environment conducive to health, well-being and learning”. This is an important responsibility, considering how influential the educational experience is, to the mental health of children and adolescents. For instance, perceived stress in school is associated with peer- and adult conflicts, perceived pressure and worries about academic achievement (Murberg & Bru, 2004). Bullying, which is experienced at least once every 14 days by 8% of Norwegian lower secondary school students (Bakken, 2019), can be detrimental to mental health (Arseneault, Bowes, & Shakoor, 2010); and lastly, high and low academic self-efficacy is associated with, respectively, low and high levels of depressive symptoms (Steca et al., 2014). Furthermore, among lower secondary school students, 10-20% are affected by mental health problems (Kieling et al., 2011). In Norway, the proportion of lower secondary school students who experience depressive symptoms increases throughout adolescence (Bakken, 2017). This applies especially for girls, of whom 27% leave lower secondary school with depressive symptoms.

Nevertheless, the current core curriculum in Norway does not have any specific strategies aiming to promote mental health-related outcomes. A new curriculum for the Norwegian school system is set to be implemented from august 2020. The curriculum will introduce a new

interdisciplinary topic called “Health and life skills”, which is meant to influence teaching in all subjects:

The school's interdisciplinary topic health and life skills shall give the pupils competence which promotes sound physical and mental health, and which provides opportunities for making responsible life choices. In the childhood and adolescent years, the development of a positive self-image and confident identity is particularly important. (The Norwegian Directorate for Education and Training, n. d.)

Despite the new curriculum’s focus on mental health, specific strategies to promote mental health in the new interdisciplinary topic, remain unclear. Considering the strong association between PA and mental health among adolescents (Kandola, Lewis, Osborn, Stubbs, & Hayes, 2020), increasing school-time PA may be a feasible and effective approach for mental health promotion. This thesis focuses on the adolescent population in Norway and evaluates the feasibility of school-based PA interventions. This thesis also evaluates whether school-based PA interventions affect mental health and elucidates aspects of the relationship between physical fitness and mental health.

2.1 A brief history of physical activity and health in schools

In Norway, school-time PA, in the form of physical education (PE), was initially intended for boys only, with the purpose of preparing them for military effort (Brattenborg & Engebretsen, 2013). From 1848, this was the purpose of PE, until the beginning of the 1900s, when what is today known as The Norwegian Olympic and Paralympic Committee and Confederation of Sports (NIF) started to influence PE to be more about sports and health. Up until today, the objectives of PE have gone through many changes with the introduction of new curriculums; however, health and mental health has consistently been part of the objectives. For instance, the curriculum of 1974 specified that the subject should

promote “harmonic growth and development” and “help students take care of their bodies” (Ministry of Church Affairs and Education, 1974). The curriculum of 1987 emphasized a holistic development for the students and mentioned mental health and physical health as two important aspects of this development (Ludvigsen et al., 2014). The curriculum also stated that PE should “provide knowledge of the body and bodily functions and assist in improving the health of individual students” (Ministry of Church Affairs and Education, 1987). In the curriculum of 1997, two of the objectives stated that students should become capable of taking care of their health and that they should develop a positive body image (Ministry of Church Affairs, Education and Research, 1996). One of the PE objectives in the 2006 curriculum, which is to be replaced in 2020, stated that: “The subject shall help pupils acquire knowledge about exercise and training, lifestyle and health, and motivate them to have an active life and continue physical training into adulthood” (The Norwegian Directorate for Education and Training, 2015). In addition to the current curriculum, in 2009, 5th to 7th graders in primary school were granted the right to regular PA, outside of PE, for the specific purpose of benefitting learning, learning environment, physical- and mental health. Furthermore, in 2012, the elective subject, Physical Activity and Health, was added to the curriculum in lower secondary school with the purpose of promoting public health and elevating PA levels (The Norwegian Directorate for Education and Training, 2012).

Although the health benefits of PA are widely accepted, and health-related objectives for PE have been present in all past curriculums, the extent to which the schools should facilitate increased PA for health-related reasons, has been debated among Norwegian scholars (Borgen, Moen, Gjølme, Løndal, & Hallås, 2018; Resaland, 2018). The opposing side argues that the main objectives of PE are that students learn about the body and with the body, and are inspired to lead a physically active life; while school-time PA lacks the learning aspect by primarily

focusing on health, and that there is little evidence to support an effect (Borgen, Gjølme, & Hallås, 2017). What both sides agree upon, however, is that more research is needed on the potential effects of additional school-time PA, in order to firmly establish what role health, PE and PA should have in Norwegian primary and secondary schools, today and in the future.

2.2 The School in Motion study - background

This thesis presents findings from the School in Motion study, a multicenter cluster randomized controlled trial (RCT) and a collaborative effort between The Norwegian School of Sports Sciences, where the lead research team was located, the University of Stavanger, the University of Agder and the Western Norway University of Applied Sciences. The study was carried out in the school year 2017/18, and the participants were 2084 9th grade students from 29 lower secondary schools.

The School in Motion study is a culmination of a series of political reports and initiatives regarding public health and the role of health-related issues in school. Already in 1946, Rolf Hofmo believed that increasing the population's participation in sports and PA, improves their health and cultural mindset, which in turn would positively impact work productivity (Goksøyr, 2007). For 20 years as the senior government official in the sports sector, Hofmo worked to incorporate into the Norwegian society that the government is responsible for facilitating PA. Years later, in 2003, a Parliament White Paper was published, titled "Prescriptions for a healthier Norway" (Ministry of Health and Care Services, 2003). It stated that physical inactivity was one of society's most concerning challenges and attributed increased population overweight and type-2 diabetes to reduced PA levels.

In 2004, the White Paper "Action plan on physical activity" (Ministry of Health and Care Services), was published as a direct response to the

challenges related to low PA levels. The main objectives stated in the action plan was to increase the amount of the population who fulfilled the PA recommendations. Although schools were mentioned as an important arena for this effort, no specific plan involving schools was specified. The action plan also aimed to increase knowledge of the population's PA levels and introduced "a system of monitoring the level of physical activity in the population" (p. 31). The results from this monitoring system raised concerns, as they showed that only 58% and 43% of 15-year-old boys and girls, respectively, met the PA recommendations (Kolle, Støren Stokke, H Hansen, & Anderssen, 2012).

In 2015, the White Paper, "The Public Health Message", was published (Ministry of Health and Care Services, 2015). It contains contrasting statements to the action plan from 2004, which emphasized that parents had the main responsibility for ensuring that children and adolescents are sufficiently active. The Public Health Message, however, based on the United Nations Convention on the Rights of the Child, stated that the government has a responsibility to facilitate sufficient PA and to protect children and adolescents from the dangers of inactivity. This White Paper also presented the initiation of a research project targeting lower secondary school, aiming to study how four hours of weekly school-time PA and PE influenced physical health, mental health and academic achievement. This was the initiation of the School in Motion study.

2.3 Contribution and aims

It has been established that PA can prevent and treat mental health problems among adults (Bennie, Teychenne, De Cocker, & Biddle, 2019; Harvey et al., 2017; Rosenbaum, Tiedemann, Sherrington, Curtis, & Ward, 2014; Stathopoulou, Powers, Berry, Smits, & Otto, 2006). Although somewhat inconclusive, recent reviews point to the same tendency among adolescents (Bailey, Hetrick, Rosenbaum, Purcell, & Parker, 2017; Biddle, Ciaccioni, Thomas, & Vergeer, 2019). Therefore,

low levels of PA and physical fitness among adolescents might have a negative influence on their mental health. Considering that the new curriculum provides schools with extended responsibilities for the students' mental health, it is important to acquire knowledge that can contribute to the fulfilment of the responsibility. This includes examining the effect and feasibility of school-based PA programs and examining the role of physical fitness to determine whether the different components have different relationships with mental health.

The majority of school-based PA intervention studies examines the effect on learning outcomes in child populations (Singh et al., 2019). Thus, whether these interventions affect mental health in average, generally healthy adolescent populations has only been investigated a handful of times (Christiansen et al., 2018; Eather, Morgan, & Lubans, 2016; Smith et al., 2018), and never in Norway. Furthermore, process evaluations regarding how school-based PA interventions are implemented are also scarce (Daly-Smith et al., 2018; Naylor et al., 2015; Watson, Timperio, Brown, Best, & Hesketh, 2017), which limits the ability to explain intervention outcomes and evaluate the feasibility of the intervention. Additionally, the roles of the different health-related components of physical fitness for adolescent mental health are scarcely explored, and the existing studies are characterized by various limitations. For instance, only one study has controlled for CRF, muscular strength and body composition (Yeatts, Martin, & Petrie, 2017), and only one study has had a longitudinal design with an average population (Ruggero, Petrie, Sheinbein, Greenleaf, & Martin, 2015).

Therefore, the primary aims of this thesis were to elucidate different aspects of the relationship between school-based PA interventions, physical fitness and mental health among adolescents, and the feasibility of school-based PA interventions in lower secondary school. The pursuit of these aims resulted in four separate research papers, each using a different methodological approach. The aims of the individual research papers were:

Paper 1: To investigate the relationship between health-related components of physical fitness and mental health in Norwegian adolescents.

Paper 2: To evaluate the implementation of intervention model 2 (The School in Motion study evaluated two separate intervention models)

Paper 3: To assess the effect of two school-based PA interventions on adolescents' mental health.

Paper 4: To investigate whether one-year changes in physical fitness components were associated with mental health status among Norwegian adolescents.

3 Theoretical framework and empirical background

This section presents the body of knowledge which the research papers of this thesis are founded upon.

3.1 Mental health

According to the annual report on Norwegian youth by Bakken (2019), the proportion of boys reporting psychological difficulties increases from 6% at the beginning of lower secondary school, to 14% at the end of upper secondary school. Among girls, the equivalent increase is from 16% to 33%. To avoid depicting adolescence as a period of illness, distress and problems, it is important to remember that various degrees of psychological difficulties are a natural part of adolescent development, and that most adolescents endure a normal development with normal mental health. However, it is also important to prevent that psychological difficulties develop into more severe mental health problems, which are likely to have several detrimental consequences. For instance, adolescent mental health problems are associated with poor academic achievement and low school attendance (DeSocio & Hootman, 2004); risky sexual behavior (Donenberg & Pao, 2005); conflicts with parents (Marmorstein & Iacono, 2004); poor general health and work impairment at age 20 (Keenan-Miller, Hammen, & Brennan, 2007). Furthermore, mental health problems occurring in adolescence can continue into adulthood (Patel, Flisher, Hetrick, & McGorry, 2007). This has economic consequences, as shown in the study by Kinge, Sælensminde, Dieleman, Vollset, and Norheim (2017), in which the findings suggested that mental disorders were the costliest conditions in Norway.

3.2 Physical activity and fitness

Keeping in mind that boys are more active than girls, the general trend for both sexes is similar: PA levels decrease throughout childhood and adolescence. The proportion of Norwegian six-, nine- and fifteen-year-olds who meet the recommended PA levels are, respectively, 91%, 82% and 50% (Dalene et al., 2018; Kolle et al., 2012). Comparing these levels with historical PA levels in Norway and internationally is difficult, because objective monitoring of Norwegian children and adolescents' PA levels did not commence until the early 2000s; hence, there are few results to compare with and the small amount of existing historical data are based on self-report measures (Samdal et al., 2007). One large population-based survey indicated that PA levels among children and adolescents remained stable between the 1980s and the early 2000s (Samdal et al., 2007). More recent international data indicate that 81% of adolescents between 11 and 17 years old did not meet the recommended PA levels, and that inactivity had increased since 2012 (Sallis et al., 2016).

There are other forms of evidence indicating that PA levels have decreased among Norwegian adolescents. A study by Dyrstad, Berg, and Tjelta (2012) compared 3000 m running performances of upper secondary school students between 1969 and 2009. Compared to the preceding decades, the mean times between 2000 and 2009 were 10% slower for boys and 6% slower for girls, indicating a decrease in CRF and aerobic PA after the year 2000. Another indication of reduced PA comes from data showing an increase in body mass index (BMI) among Norwegian adolescents between the 1960s and 1990 (Mykletun, Lydersen, Bjørnelv, & Holmen Turid, 2007). Recent data suggest that the BMI levels have stabilized (Haugen, Høigaard, & Seiler, 2014).

3.3 Physical activity, physical fitness and mental health

Regarding specific mental health outcomes in adolescent populations, studies have shown associations between vigorous PA and increased well-being (Costigan, Lubans, Lonsdale, Sanders, & del Pozo Cruz, 2019); sports participation and lower odds of later onset of depression (Easterlin, Chung, Leng, & Dudovitz, 2019); fulfilment of PA recommendations and increased mental toughness (Gerber et al., 2012); increased PA levels and increased levels of self-worth (Haugen, Säfvenbom, & Ommundsen, 2011); and between high intensity aerobic exercise and reduced stress (Norris, Carroll, & Cochrane, 1992).

Only a few studies have been conducted on the relationship between the health-related components of physical fitness and mental health outcomes among adolescents; however, existing studies show a clear trend: Kelly et al. (2010) found an association between increased CRF and reduced depression among obese females; Lubans and Cliff (2011) found an association between higher muscular strength and more self-worth among boys and girls, and between lower body fat and more self-worth among girls; Eddolls et al. (2018) found that higher CRF and lower BMI were associated with higher mental well-being and quality of life; Greenleaf, Petrie, and Martin (2010) found that lower BMI was associated with higher self-esteem, and that higher CRF was associated with higher self-esteem among girls and higher body satisfaction among boys; Padilla-Moledo, Ruiz, Ortega, Mora, and Castro-Pinero (2012) found that lower muscular strength increased the odds of reporting lower perceived health status and lower life satisfaction; and Yeatts et al. (2017) found that CRF moderated the association between neuroticism and depression. With the exception of the study by Kelly et al. (2010), all the above-mentioned studies were cross-sectional. However, the study by Kelly et al. only examined a specific sub-population, limiting generalization of results. In comparison, Ruggero et al. (2015) studied the prospective association between physical fitness at baseline and

depression one year later, in a generally healthy and average adolescent population. They concluded that higher CRF at one time point had a protective effect against the development of depression a year later.

3.3.1 *Explanatory mechanisms*

Although several studies indicate a beneficial relationship between PA, physical fitness and mental health outcomes, limited evidence exists regarding the reasons for the relationship. Lubans et al. (2016) have outlined three possible broad mechanisms that may be responsible for how PA and/or physical fitness influence mental health outcomes: neurobiological, sociocultural, and behavioral. In general, these mechanisms represent aspects of the individual that can positively or negatively affect various aspects of the individual's mental health. Moreover, which mechanisms, and how they are affected by PA, may depend on moderating factors related to the activity, such as "frequency, intensity, time, type and context" (Lubans et al., 2016). The mechanisms and how they may relate to mental health and the moderating factors, are outlined below.

3.3.1.1 The neurobiological mechanism

The neurobiological mechanism is a common term for changes in the brain that can influence mental health outcomes. The review by Matta Mello Portugal et al. (2013) presents many examples: Aerobic exercise activates the hypothalamus-pituitary-adrenal axis (also known as the stress axis), causing a protective effect against stress by reducing the stress-hormone cortisol. Dysregulation of the hypothalamus-pituitary-adrenal axis and prolonged high levels of cortisol is a hypothesized explanation of depression (Nestler et al., 2002). Exercise also influences dopamine and serotonin, which respectively are associated with the brain's reward center and anxiety levels; exercise can cause a release of endocannabinoids, which can lead to a sense of well-being; and, exercise can induce neuroplasticity, which may reduce depressive symptoms

(Matta Mello Portugal et al., 2013). Moreover, the anterior cingulate cortex (a key brain region involved in mood regulations) has been associated with mood disorders such as depression, and a recent study showed that 12 weeks of aerobic exercise produced favorable changes in the brain region (Lin et al., 2020). PA can also increase cerebral blood flow, availability of dopamine and norepinephrine, and brain derived neurotrophic factor (important part of neurodevelopmental processes), thereby positively affecting deficits that are associated with attention deficit hyperactivity disorder (ADHD; Gapin, Labban, & Etnier, 2011).

According to the review by Silverman and Deuster (2014), people with high CRF are more resilient against stress and have a lower level of C-reactive protein (a biomarker for inflammation, of which high levels are associated with depression). Moreover, CRF is associated with improved brain connectivity, thereby potentially protecting from, or treating the detrimental effect depression has on brain connectivity (Voss et al., 2016; Zhang et al., 2016). Body composition can also influence mental health outcomes through a neurobiological mechanism, as high levels of abdominal fat contributes to the production of pro-inflammatory cytokines, which can be detrimental to resilience and well-being (Silverman & Deuster, 2014).

3.3.1.2 The psychosocial mechanism

The term “psychosocial” is described by the American Psychological Association (n. d.) as “the intersection and interaction of social, cultural, and environmental influences on the mind and behavior”. Thus, the psychosocial mechanism can influence mental health through certain social, cultural and environmental interactions. An important part of adolescence is developing an identity (Ragelienė, 2016), which is made up of senses of senses of “self”, such as self-esteem, self-perception and self-efficacy (American Psyhology Association, 2002; Collins, Booth, Duncan, Fawkner, & Niven, 2019). Adolescents often evaluate themselves in light of the “cultural standard” or peers, parents and

teachers (Grabe, Ward, & Hyde, 2008; Lawler & Nixon, 2011; Schaefer & Salafia, 2014). Moreover, in adolescence, this evaluation is strongly based on appearance, since physical appearance is the most important aspect of adolescents' sense of self (Harter, 2003; Harter & Whitesell, 2001). These psychosocial mechanisms influence global self-esteem, which can predict depressive symptoms in adulthood (Steiger, Allemand, Robins, & Fend, 2014). Closely connected is social support, which can be defined as "an individual's perceptions of general support or specific supportive behaviors (available or enacted upon) from people in their social network" (Malecki & Demaray, 2002). In turn, high levels of social support can reduce the risk of depression (Turner & Brown, 2010). There are many examples that illustrate how PA or physical fitness influence psychosocial mechanisms. For instance, body image, which can be influenced by cultural norms (Grabe et al., 2008), can also be negatively influenced by high BMI (Streeter, Milhausen, & Buchholz, 2012), low CRF (Olive, Byrne, Cunningham, & Telford, 2012) and low muscular strength (Lubans & Cliff, 2011). Additionally, social support has been shown to be associated with sports participation (Babiss & Gangwisch, 2009). Social support within the sports context can be given by coaches, peers and parents, and can influence whether the athlete's sport experience is perceived positively or negatively (Sheridan, Coffee, & Lavallee, 2014).

The Basic Psychological Needs theory also proposes a psychosocial explanation for the effect of PA on mental health (Teixeira, Carraça, Markland, Silva, & Ryan, 2012). The argument is that having one's psychological needs realized (need for relatedness, competence and autonomy) contributes to well-being and that the needs can be realized when engaging in sports and exercise. Specifically, if the activity facilitates an individual experiencing social connections, mastery and a sense of autonomy, it can improve well-being. According to the theory, "…only autonomously regulated behaviors can translate into enhanced psychological wellness" (Teixeira et al., 2012)

3.3.1.3 The behavioral mechanism

Examples of behavioral mechanisms that may influence mental health outcomes are self-regulation, sleep quality and coping with stress (Lubans et al., 2016). Although the definition of self-regulation, and which components it contains is debated, it can be called "an integrative construct that includes controlling, directing, and planning cognitions, emotions, and behavior" (McClelland & Cameron, 2011, p. 32). Self-regulation has been associated with success in school and employment, which is important for well-being (Steinberg, 2014, p. 177). Furthermore, sleep disruption has been associated with psychiatric disorders (Tarokh, Saletin, & Carskadon, 2016) and improved coping skills have been associated with fewer depressive symptoms (Puskar, Sereika, & Tusaie-Mumford, 2003). Examples of PA that may influence these behavioral mechanisms include active play during recess, which has been associated with improved self-regulation (Becker, McClelland, Loprinzi, & Trost, 2014); various aerobic PA exercises, which have been associated with self-regulation skills, such as attention and concentration (Laberge, Bush, & Chagnon, 2012); accelerometer assessed PA, which has been associated with improved sleep quality (Lang et al., 2013); and sports participation, which has been associated with improved coping skills (Wijndaele et al., 2007).

CRF has also been associated with a behavioral mechanism: Stroth et al. (2009) found that adolescents with high CRF scored higher on executive control, which is part of the self-regulation construct. Interestingly, the same study also found that an acute bout of 20 minutes aerobic exercise did not influence the executive control. This suggests that the relationship between PA and self-regulation may come from average PA levels over time, rather than single bouts. In a similar study, Lott and Jensen (2016) found that CRF's association with emotion regulation in children was mediated by executive function.

3.4 Evaluating complexity

Complexity theory states that when the complexity of a context reaches a certain threshold, there is an emergence of behavior or other elements that cannot be predicted by a priori knowledge of the context (Mason, 2008). In this context, complexity, which can be a factor of the intervention, the environment or both (Hawe, 2015), can pose a challenge to accurately evaluating an intervention in real-life conditions, such as schools. RCTs are often used to evaluate interventions in schools, however, the original use for the RCT was to test the effect of a medication or other treatments in carefully controlled conditions (Craft & Editors, 1998). This is often referred to as an efficacy trial, whereas the former is referred to as an effectiveness trial (Revicki & Frank, 1999). Moreover, a school can be characterized as a complex system (Keshavarz, Nutbeam, Rowling, & Khavarpour, 2010), which, according to Shiell, Hawe, and Gold (2008) is "…adaptive to changes in its local environment, is composed of other complex systems (for example, the human body), and behaves in a non-linear fashion (change in outcome is not proportional to change in input)". Consequently, when implementing an intervention in a school with the hypothesis that it will cause a certain outcome, the system as a whole, and the smaller subsystems can adapt and interact in different ways with the intervention and influence how it is implemented (Clarke, 2010). Based on the challenges complexity poses for assessing cause and effect in effectiveness trials, Rutter et al. (2017, p. 1) have argued that:

A shift in thinking is required, away from simple, linear, causal models, to consideration of the ways in which processes and outcomes at all points within a system drive change. Instead of asking whether an intervention works to fix a problem, researchers should aim to identify if and how it contributes to reshaping a system in favorable ways.

The intervention can add further complexity to the context. A complex intervention is defined by Craig et al. (2008) to contain the following

elements: 1) many interacting components within the experimental and control interventions; 2) those delivering or receiving the intervention performs many and/or difficult behaviors; 3) many groups or organizational levels are targeted by the intervention; 4) many and potentially various outcomes; and 5) the intervention is flexible and tailoring to the specific context is permitted. According to Bonell, Fletcher, Morton, Lorenc, and Moore (2012), the usual approach when using an RCT to evaluate complex public health interventions is to ask, "what works". However, asking "what works" implies there is a "simple" intervention, which only demands that a certain dose be delivered similarly at each intervention site (Hawe, Shiell, & Riley, 2004). This is a limited understanding of causation, considering that there are many intervention components that interact dynamically with each other and the context, and influence the outcome (Hawe, Shiell, & Riley, 2009). Rather than only asking "what works", which would be sufficient in an efficacy trial, Bonell et al. (2012) advocates extending the question to "what works for whom in what circumstances". Furthermore, Deaton and Cartwright (2018) advocates the importance of also understanding "why things work", by relying on a wider set of methods in the pursuit of causal inference.

As a supplement to the RCT, a process evaluation can assist in answering these extended questions (Craig et al., 2008) by exploring "the implementation, receipt, and setting of an intervention and help in the interpretation of the outcome results" (Oakley, Strange, Bonell, Allen, & Stephenson, 2006). For instance, an intervention that does not show effect is not necessarily ineffective. Rather, the ineffective results could be due to poor implementation and without a process evaluation, researchers may dismiss their intervention, which could be effective if implemented properly. This is called a type 3 error (Dobson & Cook, 1980). On the other hand, a poorly implemented intervention can also produce beneficial results (Moore, Raisanen, Moore, Din, & Murphy,

2013); however, the point is that without a process evaluation, researchers cannot fully understand the reasons behind the outcome.

There are many definitions of implementation. Fixsen (2005) defines implementation as "a specified set of activities designed to put into practice an activity or program of known dimensions". Fullan (2016) defines implementation as "the process of putting into practice an idea, program, or set of activities and structures new to the people attempting or expecting to change". Rogers (1995) claims "Implementation occurs when an individual (or other decision-making unit) puts an innovation into use", and Stoll and Fink (1996) claims "Implementation consists of early experiences of putting reforms into practice". Durlak and Dupre (2008) presented the following definition: "In general, implementation refers to what a program consists of when it is delivered in a particular setting". Although these definitions are similar in many ways, the lack of a standard, precise definition has made it difficult to research implementation (Fixsen, 2005). This thesis uses the definition by Durlak (2016): "Implementation can be generally defined as the ways a programme is put into practice and delivered to participants". Furthermore, Durlak and Dupre (2008) identified eight aspects to describe implementation:

1) Fidelity refers to the extent to which the main components of the intervention have been adhered to.
2) Quality of delivery refers to how well the intervention components have been executed.
3) Responsiveness refers to the extent to which the participants perceive the intervention to be relevant, useful or otherwise enjoyable.
4) Dose refers to how much of the intervention that has been delivered. However, there is a distinction between dose delivered and dose received. While dose received involves the degree of actual participation and is sometimes referred to as attendance, dose delivered involves the amount of intervention that is

provided to the participants (Berkel, Mauricio, Schoenfelder, & Sandler, 2011). Dose received is not a part of the definition of implementation that is used in this thesis, but is nevertheless included, because it provides important information about participation, which can be more influential of the intervention outcomes than deliverance (Durlak & Dupre, 2008).

5) Reach refers to what extent the participating population is representative.
6) Differentiation refers to the uniqueness of the intervention, i.e. how different it is from usual practice or previous interventions.
7) Monitoring control conditions refers to the importance of knowing how different the intervention group is from the control group.
8) Adaptation refers to what extent changes have been made to the intervention in order to improve the fit in a given setting.

This thesis will primarily focus on fidelity, quality, responsiveness, dose delivered, dose received and adaptation. This focus draws on the argumentation by Berkel et al. (2011), claiming that the aspects "occur within the delivery of program sessions, and as a result, constitute potential sources of disconnect between the program as designed and that which is implemented". Moreover, fidelity, adaptation, quality and dose delivered are mainly determined by the intervention providers, while responsiveness is determined by the participants. This directs the focus at the dynamic relationship between provider and participant, who can influence each other, and ultimately, the intervention results (Berkel et al., 2011). Dose delivered and dose received were included as two separate terms because they offer information about the relationship between the provided amount and student attendance. Because implementation has such a strong influence over the outcome, the context, and how it interacts with the intervention components, implementation can be more significant for the outcome than the intervention itself (Durlak & Dupre, 2008). Evaluating the

implementation process facilitates a better understanding of the complexity of the intervention, the complexity of the context and the interaction between the two, ultimately for the purpose of improving interpretation of the outcome (Moore et al., 2015).

3.5 Previous interventions

There exist only a few studies in which school-based PA interventions have been implemented with the purpose of influencing adolescent mental health.

Bonhauser et al. (2005) conducted a 10-month PA intervention with adolescent students from an area with low socioeconomic status (SES) area in Chile and measured the effect on several mental health outcomes. Their results showed that the intervention group had decreased anxiety, increased self-esteem and unchanged depressive symptoms. Casey et al. (2014) also examined a low-SES population, however, the intervention targeted girls only, in Australia. After a one-year intervention period, the results showed that the intervention group had higher psychosocial functioning than the control group. Eather et al. (2016) conducted an 8-week intervention with Australian adolescents. There was no effect on the overall study sample, however, subgroup analyses revealed that participants with high levels of psychological difficulties at baseline showed improved self-esteem, perceived appearance, physical self-concept, and decreased psychological difficulties. Also in Australia, Lubans et al. (2016) conducted an intervention with adolescent boys from low-SES areas. After the 20-week intervention period, the results showed a small effect on psychological well-being for the overall study population. Christiansen et al. (2018) examined the effect of a 9-month intervention in a large group of Danish 10-13-year-olds. For the overall population, the results showed no effect on physical self-worth, self-perceived sports competence, body attractiveness, self-perceived social competence or global self-worth. However, interaction analyses showed that intervention group students who did not participate in leisure time

sports experienced a higher increase in self-worth, compared to their control group counterparts. Lastly, Smith et al. (2018) conducted a 10-week intervention with Australian adolescents. The intervention had no effect on self-esteem or well-being in the total population. Subgroup analyses showed effects on self-esteem in the overweight/obese subgroup. Mediation analysis also showed that self-efficacy could be an explanatory mechanism for the positive effect on self-esteem.

Among the six abovementioned studies, five used a randomized controlled design. Among those five, only three examined an general community population, while two examined single sex, low-SES groups. Moreover, only two of the studies carried out process evaluations and only one study was conducted in Europe. Although limited, the available results underline one important point: Subgroups that display low PA levels, high body composition or poor mental health at baseline are more likely to experience effects from school-based PA interventions. This is in accordance with Cerin (2010), pointing out that PA affects mental health heterogeneously, thus warranting subgroup investigations to better understand how and why PA can affect different people under different circumstances. For instance, subgroups displaying lower PA levels and poorer mental health than the population outside of the subgroups include immigrants (Abebe, Lien, & Hjelde, 2014; Sagatun, Kolle, Anderssen, Thoresen, & Søgaard, 2008), low SES populations (Bøe, Øverland, Lundervold, & Hysing, 2012; Heelan et al., 2010), girls (Bakken, 2019; Dalene et al., 2018) and poor mental health populations (Pinto Pereira, Geoffroy, & Power, 2014).

To improve the understanding of the relationship between PA and mental health-, and cognitive outcomes, many have recommended that future studies should investigate subgroups (Biddle et al., 2019; Singh et al., 2019). The use of subgroups is debated; however, adhering to certain guidelines may preserve the credibility of subgroup results (Sun et al., 2012; Wang & Ware, 2013). According to Sun et al. (2012), there are ten criteria that can be used to assess the credibility of subgroup effects.

Without listing all ten, they pertain to the design of the study, the analysis procedure and the context, e. g., congruence with previous findings. Many studies that report subgroup effects fail to meet these criteria; however, this thesis adheres to the three most critical criteria used to assess credibility of subgroup effects as they are stated by Sun et al. (2012): 1) subgroup variables must be assessed at baseline, 2) subgroup hypotheses must be specified ahead of analyses, and 3) there must be an interaction effect. The specification of expected heterogeneous effects that are outlined in the paragraph above, fulfills the second of these critical criteria.

4 Research questions

Chapter 2 has outlined the political, societal and educational background, the overarching aims of the thesis and the individual aims of each research paper. Chapter 3 has outlined the empirical background and relevant theoretical frameworks. In light of this foundation, and to guide the pursuit of the aims, this thesis poses three research questions:

1. What is the nature of the association between physical fitness components and mental health among students in Norwegian lower secondary school?
2. How is a complex school-based PA intervention implemented in Norwegian 9th grades and what influences the implementation process?
3. To what extent can two school-based PA interventions affect the mental health of Norwegian 9th graders?

5 Methods

In order to answer the research questions of this thesis, both quantitative and qualitative methods were applied. This section presents, separately, how the two methodologies were used to reach the aims of the individual research papers.

5.1 Quantitative study designs

The quantitative methods were chosen because they fit the aims of their respective papers and because they allowed the overarching topic of PA and mental health to be investigated from multiple angles. Respectively, *Papers 1, 3 and 4* used a cross-sectional-, cluster-RCT and prospective cohort design. However, all papers rely on the study population, outcome measures and procedures from the School in Motion study. Figure 1 shows the participant flow from enrollment to follow-up, missing data from each variable, and the full study population that were used in each of the quantitative papers.

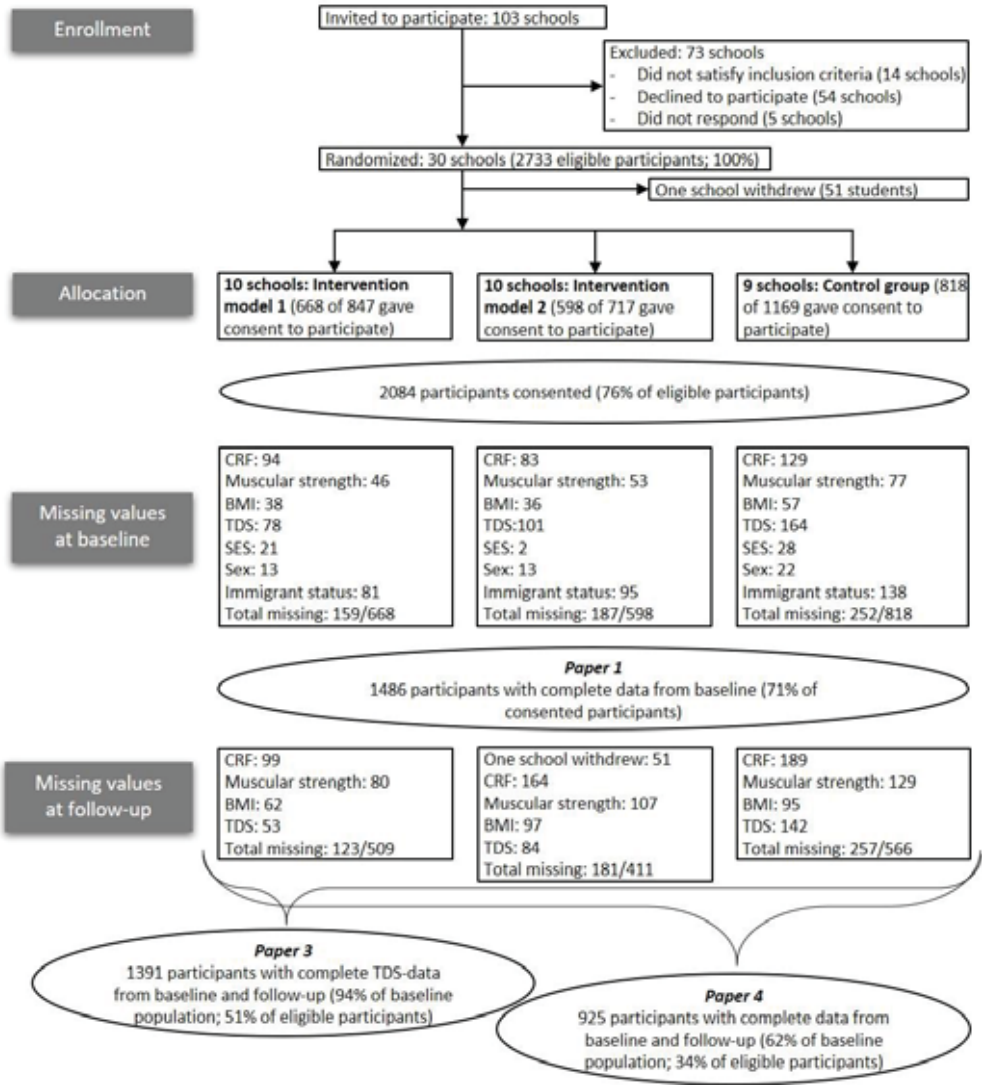


Figure 1. Participant flow, missing data and participant distribution for the quantitative research papers. CRF = cardiorespiratory fitness; BMI = body mass index; TDS = total difficulties score; SES = socioeconomic status.

5.1.1 *Participants*

The School in Motion study was a multicenter cluster-RCT and recruitment was done locally, by representatives from each test center. Thus, lower secondary schools were recruited from regions in, and surrounding Bergen, Stavanger, Kristiansand and Oslo. Private schools; special schools; schools with fewer than 25 9th grade students; and schools that already had extended PA and/or PE as part of their schedule were excluded. Out of 103 invited schools, 30 accepted to participate in School in Motion, but one school withdrew before baseline testing commenced. The final 29 schools were randomly assigned to either one of the two intervention groups, or the control group. A neutral third party was responsible for the randomization process, which resulted in the following distribution: Intervention group 1 was assigned ten schools, intervention group 2 was assigned ten schools, and the control group was assigned nine schools.

After the schools had accepted to participate, invitations to give consent for measurement participation were sent out to the parents of all 2733 eligible participants. Only students who would attend 9th grade during the intervention period were invited. Informed parental consents were obtained from 76% (n = 2084) of the potential participants. The remaining 649 students participated in the intervention but were not tested.

5.1.2 *Interventions*

The two intervention models that were evaluated in the School in Motion study were formulated by a research team containing representatives from all test centers. Initially, seven schools participated in a pilot project, in which the interventions were tested and subsequently adjusted ahead of the School in Motion study.

The interventions shared three common aspects. First, one extra PE lesson was added to the pre-existing weekly schedule. Second, the

participating schools were asked to allocate time from other subjects for the remaining intervention elements. Third, teachers at the respective schools provided the interventions, after attending mandatory training workshops. The intervention period was carried out for 29 weeks during the school year of 2017/2018 and the complete School in Motion timeline can be viewed in Figure 2.

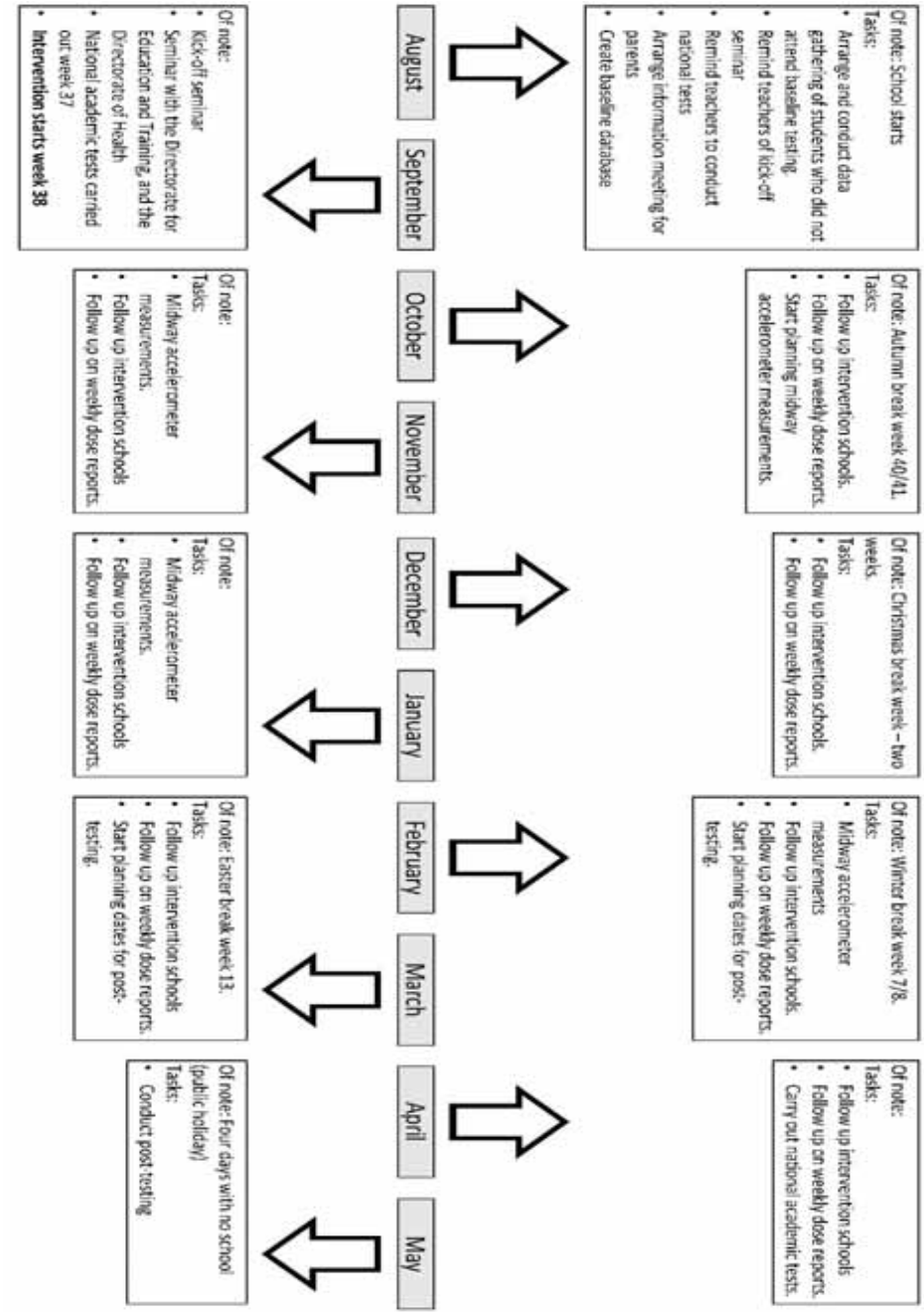


Figure 2. School in Motion project timeline

5.1.2.1 Intervention model 1

Theoretical background

Intervention model 1, "Active Learning" draws upon three important theoretical perspectives: Physical Literacy, Self-efficacy and Basic Psychological Needs Theory.

Physical literacy refers to having confidence in one's physical competence and to receive positive feedback from the experiences of one's embodied selves (Whitehead, 2010). Physical literacy is a key concept in understanding human development as a holistic process. From a holistic point of view, children develop and learn physically, mentally and socially. These areas are in continuous influence of each other, which means that learning and development in one area, may positively influence another. This is strongly connected with Bandura's (1982) theory of self-efficacy, which according to the theory, is a cognitive mechanism that mediates all behavior change. Self-efficacy means an increased belief in oneself performing successfully, which increases the willingness to attempt a task or an unfamiliar challenge (Jarvis, 2006). Sallis and Owen (1998) found positive relationships between PA and self-efficacy, self-esteem or self-concept. Consequently, increased self-efficacy can positively influence mental health outcomes, such as anxiety and depression (Muris, 2002).

The Basic psychological needs theory of motivation and behavior change has become quite popular among PA researchers in their efforts to better understand how to maximize participation and adherence to exercise programs (Vlachopoulos, Ntoumanis, & Smith, 2010). Three universal psychological needs form the basis of the theory: Autonomy reflects the need for volition in behavior; competence reflects the need for expressing and developing one's abilities; whereas relatedness reflects the need for the experience of belonging (Ryan & Deci, 2002).

According to the theory, satisfying the needs for autonomy, competence and relatedness increases the intrinsic motivation. Consequently, PA interventions that are meant to induce motivation for participation should be designed to meet these needs.

Intervention description

Active Learning consisted of three main components:

1) One additional PE lesson per week. Instructions were that the activities in this lesson should facilitate high intensity PA and possibly activities related to motor skills development. All activities were carried out in accordance with the ordinary PE curriculum.

2) Physically active academic lessons (PAAL), totaling 30 minutes per week. In PAAL, teachers in traditional classroom subjects, such as math, language, and science, use PA as a teaching method. This can be done, for instance, by having pupils run relays while they solve math problems. The instructions were that PAAL should preferably be carried out outside. Additionally, PAAL should be carried out during days when students did not have ordinary PE. The 30 minutes could be divided as the teachers saw fit.

3) One 30-minute PA lesson per week. Instructions were that teachers develop the lesson in cooperation with students, which means that students were encouraged to participate in deciding what specific activities to pursue. The focus of the lessons was supposed to be mastery, joy and self-determination. Variation throughout the intervention period was important. These lessons should preferably be carried out on days where students did not have PE or PAAL.

5.1.2.2 Intervention model 2

Theoretical background

Intervention model 2, "Don't worry, be happy" (DWBH) draws upon three theoretical perspectives: Positive Youth Development (PYD), Relational Developmental Systems (RDS) and Positive Movement Experiences (PME).

The first theory, PYD, can be seen as one of three different, but related concepts:

1. *A developmental process*
2. *A philosophy or approach to youth programming*
3. *Instances of youth programs and organizations focused on fostering the healthy or positive development of youth* (Lerner & Lerner, 2013)

According to Lerner and Lerner (2013), development is dependent on the person-context relationship. Researchers adhering to PYD suggest that youth development can be interpreted through the incorporation of the "Five Cs": competence, confidence, connection, character and caring. Those with high amounts of the "Five Cs" are likely to develop a sixth C: "Contributions to self, family, community, and to the institutions of a civil society" (Lerner & Lerner, 2013). PA interventions based on PYD entails that participants pursue PA through trial and error, self-determination, self-organization, skill-building and adult guidance, in order to achieve high levels of the Five Cs – in turn, leading to a positive and healthy adolescent development. Interventions that are based in PYD have been shown to improve well-being and prevent conduct problems and emotional distress (Taylor, Oberle, Durlak, & Weissberg, 2017).

The second theory, RDS, places its conceptual emphasis on "on mutually influential relations between individuals and contexts" (Lerner,

Hershberg, Hilliard, & Johnson, 2015). The theory considers people as active, complex adaptive systems that are creative, self-organizing and self-regulating. Furthermore, these self-driven actions take place in a socio-cultural arena; nothing happens in a vacuum. Therefore, from the perspective of the RDS, a PA intervention should not be viewed only as physical, but also mental and social. If an intervention based on RDS is expected to improve mental health, the participants need to perceive relationships within the intervention context as positive and perceive the intervention context itself as relevant. Therefore, from an RDS perspective, a PA intervention should be equally concerned with these factors as the administered amount of PA.

The third theory, PME, is an approach to understand participation in movement contexts and it attempts to create a connection between the RDS theory and PA interventions (Agans, Sävfenbom, Davis, Bowers, & Lerner, 2013). In particular, the approach argues that for interventions containing PA, exercise or sports participation to have an effect on social, physical and psychological outcomes, the intervention must provide participants with positive movement experiences. PME also opposes the belief that there is a "one-size-fits-all-intervention", thus advocating a more individualized approach. In summary, this theory claims that movement contexts in which participants have PMEs, are more likely to elicit expected effects and are more likely to increase future participation in movement contexts. PA interventions grounded in PME will be focused on creating positive and social movement contexts, with an additional focus on individual needs.

Intervention description

The primary focus of DWBH was to facilitate friendship through PA, and PA through friendship. The intervention consisted of two main components, or lessons, "Don't worry" (DW) and "Be happy" (BH). The lessons had these names to avoid being associated with traditional PE

and PA. The schools were free to choose when to conduct the two lessons but were encouraged to choose two separate days.

1) The DW lesson was supposed to be carried out as a PE lesson, but with a few adaptations: The students should be allowed to choose activities themselves and pursue them individually or in smaller groups.

2) The BH lesson had more specific instructions and deviated from ordinary PE in several ways. First, the lesson was supposed to be organized to accommodate all 9th grade classes together, because this would allow more relationships to develop. In the early stages of BH, students were supposed to think of an activity or sport that they wanted to pursue. When students knew what they wanted to do, they were supposed to find other students who also wanted to pursue the same activity, and form activity groups together. Because all of 9th grade participated together – instead of in separate classes – students could more easily find other students who wanted to pursue the same activity. It was important that teachers assisted and guided throughout this process, and especially that they could evaluate the feasibility of activities that were chosen. Next step in the process, student groups were handed planning forms, or "activity contracts". By using these, the students were supposed to make long term goals and plans for their group and their activity. They were also required to make plans for conflict resolution. The aim for the intervention was that students should be self-organized. Thus, when the activity contract was completed, the teacher approved it, and the group could commence with the activity they had planned – one lesson per week for the rest of the school year. The teacher's purpose was to guide, mentor, supervise, and provide support, but not instruct or command. The social relationships and focus on "friends in motion" were central aspects of the BH lesson.

5.1.3 *Outcome measures*

5.1.3.1 Mental health

The participants in the School in Motion study completed the Strengths and difficulties questionnaire (SDQ), which is a measure of emotional, behavioral and social components of mental functioning (Goodman, 1997). The questionnaire consists of 25 items divided into five subscales. The five subscales cover emotional problems, conduct problems, hyperactivity, peer problems and prosocial behavior. The questionnaire contains statements such as “I worry a lot”, ”I am easily distracted, I find it difficult to concentrate” and ”Other people my age generally like me”. Participants reply to the statements on a three-point Likert scale: ”not true”, ”somewhat true” and ”certainly true”. Each subscale scores from 0 to 10, and a higher score signifies a higher degree of psychological difficulties. However, a high score on the prosocial subscale signifies social strengths. The scores from the emotional problems-, conduct problems-, hyperactivity-, and peer problems subscales are summed to create the Total Difficulties Score (TDS). TDS specifically measures psychological difficulties on a scale that goes from 0 to 40, and scores can be used for assessing a general mental health state (Goodman & Goodman, 2009), as well as identifying mental disorders in a community sample of children and adolescents (Goodman, Ford, Simmons, Gatward, & Meltzer, 2000). Every one-point increase in TDS has been shown to represent a 16% to 23% increased likelihood of developing a mental disorder (Goodman & Goodman, 2009). To ease interpretation of TDS, the scale can be grouped into three levels: normal (0-15), borderline (16-19) and abnormal (20-40; Goodman, Meltzer, & Bailey, 2003). Abnormal TDS indicates a high risk of developing mental disorders. The psychometric properties of the SDQ have been validated in several countries (Goodman, 2001; Muris, Meesters, Eijkelenboom, & Vincken, 2004; Rothenberger, Becker, Erhart, Wille, & Ravens-

Sieberer, 2008), including Norway (Van Roy, Veenstra, & Clench-Aas, 2008).

The participants completed the SDQ in their own classrooms, with at least one adult present. The adult was either a teacher or part of the School in Motion test-personnel. It is important to note that the SDQ was incorporated as part of a larger questionnaire, which included other test instruments, and other general lifestyle questions. We estimated that – if filled out correctly – the full questionnaire would take 45-60 minutes to complete.

5.1.3.2 Physical fitness

Three components of health-related physical fitness were measured. CRF was measured with the Andersen-test (Andersen, Andersen, Andersen, & Anderssen, 2008). In this test, participants run between two marked lines inside a gymnasium for 15 seconds, then pause for 15 seconds, then repeat until 10 minutes have passed. The distance between the marked lines was 16 meters, and the participants were required to touch the floor behind the lines with one hand before turning and running back. A test leader kept the time and blew a whistle every 15 seconds. Other test personnel counted the number of lengths that the participants ran. According to the test protocol, the standard distance between the lines is 20 meters; however, many gymnasiums were smaller than this, which meant that the distance had to be shortened. Because of this deviation, results were not used to predict maximal oxygen consumption (Aadland, Andersen, Lerum, & Resaland, 2017); instead, the distance covered in meters was used as the outcome variable.

Muscular strength was assessed by conducting three specific exercises, as they are described in the EUROFIT test battery (Council of Europe, 1988): sit-ups, standing broad jump and handgrip test. The sit-ups involved participants attempting to do as many as they could in 30 seconds. The participants did the exercise with their knees in a 90-degree

angle and their fingers locked behind their head, and their feet held to the floor by test personnel. To get a valid count, the participants had to touch their knees with their elbows, going up, and touch the floor with their shoulders, going down. The participants had two attempts on the standing broad jump, and the best attempt was the valid outcome. Participants did the exercise by jumping as far as they could from a stand still position, and the distance was recorded from the heel closest to the starting point. Measurements were recorded to the closest cm. Lastly, participants performed the handgrip test by gripping a Baseline dynamometer (Baseline® Hydraulic Hand Dynamometer, Elmsford, NY, USA) with their dominant hand, as hard as they could for three seconds. The working arm was supposed to be held in a normal position, straight down, alongside the body. The better of two attempts was the valid result, and the results were recorded to the closest kg.

Body composition was measured by calculating individual body mass index (BMI) scores (kg/m^2) from the participants' height and weight. Participants' height was measured with a portable stadiometer (Seca 123, Hamburg, Germany). The results were recorded to the closest mm. Participants were asked to remove their shoes to be weighed. Weight was measured with a digital scale (Seca 899, Hamburg, Germany), and all measurements were recorded to the closest 0.1 kg. We asked the participants to only wear shorts or tights with a t-shirt as they were being weighed, but some wore sweaters and pants. Their clothes were noted, and their weight was adjusted in the following analysis: 1 kg was subtracted for pants and/or sweater, 0.5 kg was subtracted for shorts/tights and t-shirt.

Dose delivered

Dose delivered is expressed as the mean percentage of intervention components that were executed relative to the total number of intervention components that were possible to execute during the 29-week intervention period. One teacher liaison from each school was

responsible for reporting the intervention components as executed/not executed on an online platform.

Covariates and subgroups

Three additional categorical variables were included in analyses, either as covariates, or moderators and subgroups. Participants' sex was noted by test-personnel. Participants' immigrant status was determined by a question in the questionnaire, "were you born in Norway?", to which the participants answered yes/no. SES is expressed as parents' education level (Erola, Jalonen, & Lehti, 2016), obtained from Statistics Norway. The education levels were divided in four categories: lower secondary school or less, upper secondary school, less than four years university education, more than four years university education. If two parents had valid data, the parent with the highest level of education was included in the analyses. Assessing these potential subgroups at baseline fulfills the first of the critical criteria of credibility for subgroup results (Sun et al., 2012).

5.1.4 *Statistical analyses (Papers 1, 3 and 4)*

All data management and statistical analyses were performed in IBM SPSS Statistics 25 (IBM, Armonk, New York, USA). SDQ data were managed and organized into the predetermined scales by the syntax provided by the SDQ information web page (Youthinmind, 2018). Cronbach's alpha was employed to assess the internal consistency of TDS and its subscales. The respective alphas from baseline and follow-up were as follows: emotional problems (0.67 and 0.71), conduct problems (0.51 and 0.53), hyperactivity (0.66 and 0.68), peer problems (0.61 and 0.61) and TDS (0.62 and 0.61).

Descriptive statistics were calculated for all papers and are presented as means and standard deviations (SD). Linear mixed effects models were the main analysis used to assess associations and effects. Although

reporting practices are non-conform, this procedure is among the most widely used in psychological science (Meteyard & Davies, 2020). Linear mixed effects models are advantageous as they allow the user to apply advanced regression models to longitudinal data, and they are appropriate when data are organized in hierarchical structures (West, 2009). An example of a hierarchical structured data is in a cluster-RCT in which participants are clustered within schools. Thus, in the linear mixed effects models, the 29 participating schools were treated as random effects. Statistical procedures that were specific to each paper are outlined below.

Paper 1 – The cross-sectional association between physical fitness and mental health

Z-scores stratified for sex and BMI quartiles were created for handgrip strength, standing broad jump and sit-ups. The z-scores were used to create one composite mean z-score for muscular strength. The association between each health-related component of physical fitness and mental health was assessed separately and controlling for each other. The association between the individual muscular strength variables and mental health was also assessed separately. Follow-up TDS was the dependent variable, and all models controlled for baseline TDS, sex, immigrant status and SES. The reported regression estimates and their 95% confidence intervals (CI) reflect the change in TDS as a result of one unit of measurement change in the independent variables.

Paper 3 – The interventions' effects on mental health

Baseline differences between intervention group 1 and control, and between intervention group 2 and control, were assessed with one-way ANOVA with Fisher's LSD post hoc test. The effect of group (intervention group 1, intervention group 2 and control) on change in the dependent variables (follow-up mental health outcomes) was tested. The models included the respective baseline variable as a covariate (e. g., when the dependent variable was follow-up TDS, the independent

variable was baseline TDS). Moderating effects for change in the dependent variables were determined by testing a categorical subgroup * group interaction in models controlling for main effects on group and subgroup. The moderating variables were sex, SES, immigrant status and TDS level at baseline (normal, borderline and abnormal). By performing the moderator analyses and assessing interaction effects, the third of the critical credibility criteria for subgroup effects, set by Sun et al. (2012), was fulfilled.

When analyses indicated interaction effects, subsequent subgroup analyses were conducted with the dataset stratified by the categorical subgroup variable. The main dependent variable was follow-up TDS, and when results were interpreted to be compatible with an effect, subsequent analyses were conducted with the TDS subscales as dependent variables. The effect is reported as the estimated mean difference in change between groups (b), 95% CI and exact p-values. The difference in change is expressed by measurement units on the scale of the dependent variable, adjusted for potential baseline differences (intervention group 1 - control and intervention group 2 - control). The intraclass correlation coefficient (ICC) for the cluster effect of schools is also reported.

Paper 4 – The association between a one-year change in physical fitness and mental health status

Change scores were calculated for the physical fitness variables (follow-up - baseline) and functioned as the independent variables. The muscular strength variable is a mean composite z-score, stratified for sex, which was calculated by transforming change scores for each muscular strength component to z-scores and averaging them to one mean score. Paired t-tests were conducted with all outcome variables to assess the overall changes from baseline to follow-up. The prospective association between change in physical fitness and mental health status was primarily examined by including follow-up TDS as the main dependent

variable, while the independent variables were scores representing one-year changes in CRF, muscular strength and BMI. All models also controlled for the dependent variable's baseline result, sex, SES, immigration status and experimental group allocation (intervention group 1, 2 or control). The moderating effect of sex, SES, immigrant status, TDS at baseline and experimental group allocation was examined in separate models for each moderator (in fulfillment of the third credibility criterion by Sun et al. (2012)). In each model, the moderator was added as an interaction term for the three physical fitness variables, while also controlling for the same variables. When the results indicated an interaction effect, subgroup analyses of the categorical variables were conducted. When results indicated compatibility with an association with follow-up TDS, subsequent analyses were conducted with the TDS subscales as dependent variables (follow-up emotional problems, conduct problems, hyperactivity and peer problems – while controlling for the respective baseline variable in each model). Results are reported as unstandardized regression coefficients (b) with 95% CI and exact p-values. The b represents difference on the scale of the dependent variable for each unit increase on the scale of the independent variable.

5.1.4.1 Missing data

Complete case analysis was chosen as the main way of handling missing data. Due to a large amount of missing values, it was important to also thoroughly examine whether they caused bias (Altman & Bland, 2007), therefore extensive missing value analyses were conducted.

All three papers assessed whether missing values were missing completely at random (MCAR) with Little's MCAR test. In all three papers, grouping variables were created, separating the complete case groups from participants with at least one missing value. One-way ANOVAs were conducted to analyze the difference between the included complete case groups and the excluded missing values-groups. In *Paper 1*, Pearson's correlation analysis was also used to compare the

correlations in both groups. In *Papers 1 and 4*, descriptive and frequency statistics determined the degree of missingness for each variable. In *Paper 3*, a logistic regression model examined whether the outcome variable or any of the covariates could predict the likelihood of being a complete case or having at least one missing TDS value. The same logistic regression model examined whether baseline TDS could predict the likelihood of being a complete case. To assess whether variables predicted the likelihood of missingness differently between intervention group 1, 2 and control, the same logistic regression was carried out, stratified by experimental group. Additionally, ANOVA with Fisher's LSD post hoc test was used to assess whether baseline TDS, among those with missing follow-up TDS, was different between intervention group 1, 2 and control.

As a final measure of evaluating the missing data, multiple imputation was conducted on the datasets specific to each research paper. The three quantitative papers report results from both complete case analyses and after multiple imputation, which has been recommended by Manly and Wells (2015) and Sterne et al. (2009). However, in the discussion sections of the papers and in this thesis, emphasis is placed on the complete case results. The multiple imputation results functions as a means to elucidate that the data and results are uncertain.

5.1.4.2 Statistical output interpretation

For *Paper 1*, the p-value interpretation was based on the traditional understanding that $p < .05$ indicates a statistically significant association. In *Papers 3 and 4*, however, p-values were not dichotomously interpreted as significant or non-significant. Instead, interpretations adhered to the guidelines provided by the American Statistical Association's statement on statistical significance and p-values (Wasserstein & Lazar, 2016), and more recent guidelines in a special issue of The American Statistician (Wasserstein, Schirm, & Lazar, 2019). Therefore, p-values were interpreted as continuous quantities that

express how compatible the observed data are with the null-hypotheses: Higher p-values indicate greater compatibility with the null-hypotheses. The results in papers 3 and 4 are presented based on a continuous interpretation of the p-values, the size of the unstandardized regression estimates (b) and the limits of the confidence intervals (Amrhein, Greenland, & McShane, 2019; Greenland et al., 2016). The interpretations are also influenced by prior evidence, plausibility of mechanism, study design and data quality (McShane, Gal, Gelman, Robert, & Tackett, 2019).

5.2 Qualitative study design

The process evaluation used a cross-sectional design to gather qualitative data by conducting interviews and observations. The qualitative methods were chosen for the process evaluation because quantitative methods cannot capture specific details describing how the intervention was implemented and received. The following sections describe the qualitative approach used to collect data for *Paper 2*.

5.2.1 *Participants*

Four schools assigned DWBH, one from each of the four regions in Norway where the intervention was carried out, were randomly selected and asked to participate. The teacher liaisons at the schools were asked to perform a purposive sampling of students to be interviewed: three activity groups representing different activities and opinions toward the intervention and PA in general. The purposive sampling strategy was employed to cover a diversity of experience, in order to prevent bias toward presenting only one type of information from one type of participants (Robinson, 2014). Teachers who supervised DWBH were also interviewed to obtain knowledge about the implementation process from the providers' perspectives. A total of 54 individuals were interviewed. This amounted to 12 student focus group interviews (n =

46), two individual teacher interviews and two teacher focus group interviews (n = 6).

5.2.2 *Data collection*

The interviews were mainly semi-structured in focus groups. Two of the teacher interviews, however, were individual because of illness. Semi-structured interviews were chosen based on expectations that there could be highly varied opinions and experiences of DWBH and other emerging issues (Barriball & While, 1994). The focus group approach was chosen because it facilitates answering questions and discussing a specific topic in a social context, which potentially reveals information that would not have emerged in an individual interview (Frey & Fontana, 1991). Focus groups are therefore suitable for program evaluation. The interviews were meant to capture participants' and providers' experiences of the intervention. The interview guide was constructed to elicit answers that could be linked to the six implementation aspects: fidelity, quality, dose delivered, dose received, responsiveness and adaptation.

Observations were also conducted to experience a physical presence and to get an impression of the environmental surroundings, the participants, attitudes toward the intervention and the dynamics between providers and participants. One BH lesson in each of the selected schools were observed, by the same observer. These secondary data were gathered to improve the ability to interpret the interviews and explain causal mechanisms (Manzano, 2016). Notes from the observations assisted in the interpretation and coding of the interview material, in particular by supporting the coherence between what the interviewees stated and what was observed (Mays & Pope, 1995).

The qualitative data were gathered within the same week in each school, during the second half of the intervention period (between mid-January and mid-March 2018). The interviews took place during school hours at the participants' respective schools, in a classroom with only the

researcher and the interviewee(s) present. The interviews lasted between 30 and 55 minutes and were audio recorded.

It should be noted that dose delivered, a quantitative form of data, was reported in *Paper 2*, and was discussed in connection with the other implementation aspects. Dose delivered is expressed as the percentage of lessons provided to the participants, relative to the number of lessons that were possible to provide during the intervention period. To measure dose delivered, the teacher liaisons used an online registration tool to weekly register DWBH lessons as executed/not executed.

5.2.3 *Qualitative analysis (Paper 2)*

Audio recordings of the interviews were imported into Nvivo qualitative data analysis Software 12 (QSR International Pty Ltd., Doncaster, Australia) and transcribed verbatim. Data were further analyzed in Nvivo and Excel, by using the five steps of the framework analysis: familiarization, identifying thematic frameworks, indexing, charting, and mapping and interpreting (Spencer & Ritchie, 2002). The framework analysis was chosen because it exists within the family of content- or thematic analysis, and can be used to "identify commonalities and differences (…) focusing on relationships between different parts of the data, thereby seeking to draw descriptive and/or explanatory conclusions" (Gale et al., 2013). Briefly, one deductive and one inductive analysis were conducted consecutively. The deductive analysis focused on exploring how the intervention was implemented and the inductive analysis focused on identifying the factors that influenced implementation. The inductive analysis resulted in the merging of identified subcategories into four main factors: 1) frame factors, 2) intervention characteristics, 3) participant characteristics and 4) provider characteristics. The fifth and final step of the analysis involved combining the inductive and deductive findings to interpret and outline processes behind the implementation and the influencing factors. This

process amounted to the results, which describe how and why specific factors influenced specific aspects of implementation.

5.3 Ethical considerations

The term research ethics refers to a wide variety of values, norms, and institutional arrangements that help constitute and regulate scientific activities. Research ethics is a codification of scientific morality in practice. Guidelines for research ethics specify the basic norms and values of the research community. They are based on general ethics of science, just as general ethics is based on the morality of society at large. (The Norwegian National Research Ethics Committees, 2016)

5.3.1 *School in Motion*

All test procedures for the School in Motion study, including the interviews and observations, were approved by the Norwegian Centre for Research Data (project number 49094), and the project was in accordance with the Declaration of Helsinki for experiments involving humans. A great deal of ethical considerations went into the planning, execution and dissemination of the School in Motion study.

The participating researchers were cognizant of the potential ethical challenges regarding the use of under-age participants. Some might question doing research on young adolescents, because they are a vulnerable population. Rhodes (2010), however, argued that categorizing groups as vulnerable is paternalism and a threat to their autonomy. The label of vulnerability comes from the assumption that their ability to appraise risks and benefits is inadequate. Furthermore, the School in Motion study was relevant for the participating adolescents, therefore they were the participants (Smyth & Weindling, 1999). However, participation was voluntary, and several measures were taken to protect the volition of the involved parties: Detailed information was provided to the students and parents ahead of the intervention, and

written parental consent was required for students to participate in testing. Another consideration was to ensure anonymity and prevent the possibility of identifying participants or schools. Therefore, every participant was paired with a unique identification number, which they wore as a sticker on their sweater or t-shirt during testing. The computer file containing the paired identification numbers and names was password protected and could only be accessed by authorized test personnel. All test personnel also signed a non-disclosure agreement, agreeing not to spread any information to a third party. All collected data from the School in Motion study will be destroyed after all results have been disseminated.

One of the questions Foster (1995) recommended research committees ask when approving studies was "Are the risks to the research subjects acceptable?". No risks were anticipated from participating in the interventions, and precautions were taken to ensure that potential risks from testing procedures were kept at a minimum: For instance, adolescents can feel uncomfortable about a stranger measuring their weight. This was handled by performing these measurements in a closed room, away from other classmates and by having male test personnel measure the boys while female test personnel measured the girls. Test personnel ensured participants that being measured was completely voluntary and that the participants could opt out at any moment. Furthermore, no test results were disclosed to the participants, in order to prevent them from comparing numbers with classmates. Smyth and Weindling (1999) wrote, "The counterpoint to risks are the benefits, both to those participating in the research and to other children who may be future beneficiaries". The School in Motion study is intended to benefit its participants and future students who may receive additional curricular PA as a result of the study, and it can therefore be claimed that these benefits outweigh potential risks.

A key ethical issue with the RCT design has to do with the control group that does not receive the benefits of the intervention (Nardini, 2014).

However, to make amends for the potential disadvantages of not receiving additional school-time PA, the control schools were offered the opportunity to carry out one of the interventions the year after. These schools also received the same economic compensation as the intervention schools, so they were able to add lessons to their schedule.

"Schools are encouraged, and in some cases required, to engage with and generate research, and this is welcomed as a means by which teachers are professionally developed and learners taught in research-rich environments" (Bryan & Burstow, 2018). However, when conducting research in schools, it is important to keep in mind that the schools serve many other purposes than being a research arena. The teachers have other tasks than carrying out our intervention and the students have other challenges that will not be solved by increasing PA levels. The ethical consideration in this regard was to act humble, treat the school faculty with respect and keep in mind that researchers are guests who should be grateful for being allowed inside.

The School in Motion study was commissioned by the Norwegian Directorate of Health and the Directorate of Education and Training. The study received an unusually high amount of funding, because the participating schools were economically compensated for the extra work that followed the intervention. With this kind of funding follows certain expectations of positive results, which is followed by a pressure to deliver said results. Therefore, results that show no effect can be disappointing and might lead to the fabrication of data to make the results more presentable. For instance, in a study on doctoral students in Norway, "38% did not agree with the statement that it is never appropriate to try a variety of different methods of analysis until one is found that yields a result that is statistically significant" (Hofmann, Myhr, & Holm, 2013). However, fabrication of results from the School in Motion study would not likely go unnoticed, considering the number of researchers who are part of the study, and have access to the same data material. According to Hamilton (1990), more than 55% of published

papers are not considered worth citing by others. Gerber (2006) discusses that this may lead to many being tempted to publish fabricated results, in the belief that no one are likely to read the paper anyway. Considering the societal importance and relevance of School in Motion, many readers and much attention is expected, thus if any of the connected authors were to publish fraud papers, the threshold for detection would be extremely low. The author of this thesis is cognizant of the responsibility and potential scrutiny that comes with publishing results from a study of this magnitude. Therefore, the author of this thesis strives to ensure that the results are communicated as openly and honestly as possible.

5.3.2 *The process evaluation*

There are many ethical issues to consider when interviewing and observing adolescents. For instance, it is important to preserve the privacy and anonymity of participants (The Norwegian National Research Ethics Committees, 2016). For the observations, this was handled by only taking notes, instead of recording audiovisual material, which would require safe data storage. Additionally, the observer did not know the identities of the observed students. The groups who were interviewed are likely unidentifiable by the public, but because of statements regarding their specific schools, they might be identified by their peers or their teachers (Allmark et al., 2009). Therefore, when writing up the research paper containing the process evaluation, the authors were careful to omit details that would disclose the identity of the students. Furthermore, presentation of results can be an ethical challenge because, in some interviews, schools and teachers were sometimes described in a negative manner. It was therefore important to be respectful and balanced when presenting negative characterizations and only presenting characteristics that were relevant for the context and the aim of the process evaluation.

According to Sim and Waterfield (2019) the group setting poses a set of ethical challenges. For instance, there is a possibility that one of the

interviewees can communicate statements to someone outside of the group. This could have been a potential risk, which would have been impossible to prevent. However, the author of this thesis considers that to be an unlikely scenario. Another ethical challenge is having adolescents participate in a group discussion, which some may find uncomfortable. A standard statement was read ahead of the interview, to inform the interviewees of the interview procedure, their option to decline to respond or to opt out at any time.

The power structure of the interview was also an ethical consideration ahead of, during and after the interviews. Challenges related to the potential power that the interviewer may have over the interviewees are discussed by Allmark et al. (2009). In this case, the power relations between interviewees and interviewer may be insignificant. However, there may be a challenge related to the power relations between the teacher and the student interviewees. For instance, although participation in the interviews was voluntary, students may have felt pressured to participate when asked by their teachers.

Lastly, the importance of validity should be highlighted. In quantitative research, validity refers to the relevance of the measuring instrument and whether the researcher is measuring what the researcher intends to measure (Hammersley, 1987). In qualitative research, however, validity refers to the accurate representation of the "features of the phenomena" that an account intends to "describe, explain or theorise" (Hammersley, 1992, p. 69). Based on this definition, one can say that validity has an ethical dimension, as it entails an honest and correct representation of a phenomenon and people within it. Providing accurate representations was a tenet throughout the process of planning, executing and interpreting the qualitative data collection. This required a reflexive strategy, which can be defined as a "continual evaluation of subjective responses, intersubjective dynamics, and the research process itself" (Finlay, 2002, p. 532). However, to paraphrase Whittemore, Chase, and Mandle (2001), validity cannot be assumed by the researcher, readers

must be able to critically evaluate findings; explaining how the findings were produced is equally important as the findings themselves. With this in mind, the author of this thesis has attempted to provide a valid account of the implementation process of intervention model 2.

6 Results

This section presents the title, aim and results of each paper, separately.

6.1 Paper 1

The association between physical fitness and mental health in Norwegian adolescents

The aim of the paper was to investigate the cross-sectional association between health-related components of physical fitness and mental health in Norwegian adolescents.

This study used data from the baseline measurements of the School in Motion study. Due to much missing data (Figure 1), only 1486 participants were included in the analyses. When separately testing the associations between the physical fitness components and TDS, the results did not indicate an association for BMI (b = 0.058; 95% CI = -0.027 to 0.142; p = .184), but there were inverse associations for muscular strength (b = -0.458; 95% CI = -0.810 to -0.109; p = .010) and CRF (b = -0.006; 95% CI = -0.009 to -0.003; p < .001). The fully adjusted model, however, showed no association between muscular strength and TDS, or between BMI and TDS. The fully adjusted model showed results compatible with an inverse association between CRF and TDS (b = -0.006; 95% CI = -0.009 to -0.002; p = .001). The association between CRF and TDS suggested that every 100 m distance increase in the CRF test was associated with 0.6 points (1.5%) lower levels of psychological difficulties. Relative to the TDS-levels of the overall population (10.3), the association between 100 m distance increase in the CRF and 0.6 points lower TDS corresponds to 6% lower TDS.

6.2 Paper 2

A process evaluation of a school-based physical activity intervention: Influencing factors and potential consequences of implementation

The aim of the study was to evaluate the implementation of the DWBH intervention, by examining 1) how DWBH was implemented and 2) what influenced implementation.

The main factors that influenced the implementation were frame factors, intervention characteristics, participant characteristics and provider characteristics. The results showed large differences between the four investigated schools for how DWBH was implemented. School 2 had many students, inadequate facilities and too few teachers to supervise students who were off school grounds (frame factors → fidelity and quality). The school also added the lessons on top of the existing schedule, which caused the students to miss their school bus, which, in turn made the students dislike the intervention (frame factors → adaptations → responsiveness). Because of the negative response, many reported that truancy was common (responsiveness, participant characteristics → dose received). School 3 also had few facilities, but adaptations were made to the intervention so that the lessons were carried out in separate classes (frame factors → adaptations). Due to this scheduling, however, one of the classes did not have access to the gymnasium because another class had PE lessons (frame factors → dose delivered and dose received). They were also often without a teacher because of long-term sick leave (provider characteristics and frame factors → quality, dose delivered and dose received). Students in School 3 were still highly positive toward the intervention, which was also the case among the students in Schools 1 and 4 (intervention characteristics → responsiveness). Additionally, Schools 1 and 4 had ample facilities and made small adaptations that were advantageous for the implementation as a whole (frame factors, provider characteristics → adaptation, fidelity, quality and responsiveness). One example of an

adaptation that was positively received by the students, was made in Schools 1, 3 and 4: The supervising teachers decided that both lessons should allow students to pursue the activity of their interest, within their activity group. Thus, both lessons followed the BH guidelines. The implementation process in Schools 1 and 4 can be characterized as successful.

6.3 Paper 3

Effects of school-based physical activity interventions on mental health in adolescents: The School in Motion cluster randomized controlled trial

The study aimed to investigate whether the two PA-interventions used in the School in Motion study had any effects on participants' mental health.

Due to missing values, the analyses included 523 participants from intervention group 1, 381 participants from intervention group 2 and 487 participants from the control group (total n = 1391; see Figure 1). Results were not compatible with effects on the overall population. Interaction effects warranted subgroup analyses of immigrant status (p = .061) and baseline TDS levels (p = .008). In the abnormal TDS subgroup (n = 71), results were compatible with a mean difference in change for TDS in favor of intervention model 1, compared to their control group counterparts (b = -2.9; 95% CI = -5.73 to -0.07; p = .045). Relative to the estimated baseline levels within the abnormal TDS subgroup (23 points), intervention model 1 reduced TDS by 22%, while the control condition reduced TDS by 9%. Follow-up analyses of the TDS subscales showed that the result could mainly be attributed to difference in change for conduct problems (b = -0.99; 95% CI = -2.02 to 0.04; p = .058) and hyperactivity (b = -1.13; 95% CI = -2.1 to -0.19; p = .019).

In the immigrant subgroup (n = 118), results were compatible with a mean difference in change for TDS in favor of intervention model 1 (b =

-1.6; -3.53 to 0.27; p = .093) and intervention model 2 (b = -2.1; -4.36 to 0.21; p = .075), compared to their control group counterparts. Relative to the estimated baseline levels within the immigrant subgroup (11 points), TDS increased 5% in intervention group 1, 0% in intervention group 2 and 18% in the control group. Follow-up analyses of the TDS subscales showed that the result from intervention model 1 could mainly be attributed to difference in change for emotional problems (b = -1.1; 95% CI = -1.89 to -0.29; p = .008), and the result from intervention model 2 could mainly be attributed to difference in change for emotional problems (b = -1.0; 95% CI = -1.99 to -0.07; p = .036) and hyperactivity (b = -0.94; 95% CI = -1.90 to 0.02; p = .055).

The analyses on the immigrant/non-immigrant subgroups with the SDQ subscales as dependent variables showed an unexpected result: in the non-immigrant subgroup, there was compatibility with a mean difference in change for peer problems indicating an increase from intervention model 2 (b = 0.32; 95% CI = 0.03 to 0.62; p = .034), compared to the corresponding control subgroup. This warranted a thorough exploration of the potential heterogeneous effects on peer problems. The results were compatible with a mean difference in change for peer problems in disfavor of intervention group 2 within the non-immigrant girls subgroup (n = 652; b = 0.42; 0.11 to 0.72; p = .010) and within the borderline TDS subgroup (n = 138; b = 0.89; 0.10 to 1.67; p = .029), compared to their control group counterparts. Relative to the estimated baseline levels within the non-immigrant girls subgroup (1.6 points), peer problems increased 19% in intervention group 2, while their control group counterparts displayed a 13% decrease. Similarly, relative to estimated baseline levels within the borderline TDS subgroup, (3.2 points), peer problems increased 6% in intervention group 2, while their control counterparts displayed a 31% decrease.

6.4 Paper 4

The association between a one-year change in physical fitness and mental health status in Norwegian adolescents

The aim of the study was to investigate whether one-year changes in physical fitness components were associated with mental health status among Norwegian adolescents.

The study treated the randomized study sample as one cohort, and out of the full study sample (n = 2084), 925 participants were included, due to missing values (Figure 1). The results for the overall population were not compatible with an association between follow-up TDS and change in BMI or change in muscular strength. The results were compatible with an inverse association between change in CRF and follow-up TDS (b = -0.004; 95% CI = -0.008 to -0.001; p = .040). Interaction effects warranted subgroup analyses of sex (p = .010), immigrant status (p = .110) and SES (p = .011). Results after stratifying for sex were compatible with an inverse association between change in CRF and follow-up TDS among boys (b = -0.009; 95% CI = -0.015 to -0.003; p = .006). The results suggest that 100 m increase in CRF was associated with 8.5% (0.9 points) lower follow-up TDS, relative to boys' mean values. (10.5 points). Analyses of the TDS subscales showed that the results could mainly be attributed to lower conduct problems (b = -0.002; 95% CI = -0.005 to -0.001; p = .034), hyperactivity (b = -0.003; 95% CI = -0.006 to -0.001; p = .011) and peer problems (b = -0.003; 95% CI = -0.005 to -0.001; p = .005).

Results after stratifying for immigrant status were compatible with an inverse association between change in muscular strength and follow-up TDS among immigrants (b = -1.96; 95% CI = -4.03 to 0.092; p = .061). These results suggest that one standard deviation increase in the z-score composite of handgrip strength, standing broad jump and sit-ups, was associated with 18% (1.96 points) lower follow-up TDS, relative to the immigrant population's mean values (10.7). Analyses of the TDS

subscales showed that the results could mainly be attributed to lower conduct problems (b = -0.48; 95% CI = -1.05 to 0.08; p = .095), and hyperactivity (b = -0.64; 95% CI = -1.45 to 0.16; p = .113).

Results after stratifying for SES were compatible with inverse associations between change in CRF and follow-up TDS in SES groups 1-3 (unstandardized coefficients between -0.007 and -0.04, p-values between .039 and .063). SES group 4 (highest levels of parental education) demonstrated an opposite tendency (b = 0.004; 95% CI = -0.004 to 0.012; p = .278). This warranted further investigation of SES group 4. Results after stratifying for SES and sex were compatible with an association between change in CRF and TDS status among girls in SES group 4 (n = 135; b = 0.014; 95% CI = 0.003 to 0.025; p = .014). These results suggest that 100m increase in CRF was associated with 13% (1.4 point) higher follow-up TDS, relative to girls in SES group 4's mean values at follow-up (10.5). Subscale analyses showed that these results could mainly be attributed to more emotional problems (b = 0.005; 95% CI = -0.001 to 0.01; p = .076) and hyperactivity (b = 0.007; 95% CI = 0.003 to 0.012; p = .002).

7 Discussion

The overarching aims of this thesis were to elucidate various aspects of the relationship between school-based PA interventions, physical fitness and mental health among adolescents, and the feasibility of school-based PA interventions in lower secondary school. In pursuit of these aims, four research papers have been produced and presented as the main content of this thesis. Each paper has its distinct aim, methodological approach and results that contribute to a more solid knowledge base for the research field. In this chapter, the results of each paper are discussed and interpreted in light of the theoretical framework and empirical background, outlined in Chapter 3. The discussion is structured around the three research questions of this thesis, presented in Chapter 4.

7.1 Research question 1

What is the nature of the association between physical fitness components and mental health among students in Norwegian lower secondary school?

This question is elucidated by *Papers 1 and 4,* which respectively, aimed to investigate the cross-sectional and the prospective association between physical fitness and mental health. The main findings in *Paper 1* indicated that only CRF was inversely associated with psychological difficulties. Muscular strength, independently, was inversely associated with psychological difficulties, but not when controlling for CRF. There were three main findings in *Paper 4:* 1) a one-year change in CRF was inversely associated with follow-up psychological difficulties among boys; 2) a one-year change in muscular strength was inversely associated with follow-up psychological difficulties among immigrants; and 3) a one-year change in CRF was associated with follow-up psychological difficulties among girls in the highest SES-group.

7.1.1 *Cardiorespiratory fitness*

The findings in *Paper 1* are nearly in line with the findings of a recent study by Janssen et al., (2020), which found that CRF was the only health-related component of fitness associated with internalizing problems, which is composed of emotional problems and peer problems of the SDQ. Also similar, CRF has been found to be the only fitness component associated with higher quality of life (Andersen et al., 2017), lower depression levels (Rieck, Jackson, Martin, Petrie, & Greenleaf, 2013; Yeatts et al., 2017) and improved well-being (Kelly et al., 2010). Studies that have found associations between muscular strength and mental health outcomes in adolescents did not control for CRF (Lubans & Cliff, 2011; Padilla-Moledo et al., 2012). Adjusting for CRF has also been omitted in studies that found associations between body composition and mental health outcomes, as shown in the review of Luppino et al. (2010).

The findings in *Paper 4* add nuance to the cross-sectional relationship between CRF and psychological difficulties. Firstly, the association between a one-year change in CRF and mental health status among boys was larger (b = 0.009) than the baseline cross-sectional association between CRF and mental health in the overall population (b = 0.006). Although the unstandardized coefficients seem small, it is important to note that they represent change on the TDS-scale for every one-meter increase in the CRF-test. Considering that the overall results for the CRF-test at baseline and follow-up, respectively, were 906 m and 927 m, discussing an unstandardized coefficient based on a difference of one meter is not very meaningful. Instead, by discussing the outcomes' associations with 100 m increases on the CRF-test, the unstandardized coefficients from *Papers 1 and 4* (0.6 and 0.9) can be discussed in more practical terms. It has been suggested by Goodman and Goodman (2009) that every one-point increase in TDS may represent a 16% to 23% increased likelihood of developing a mental disorder. Thus, both associations may be of clinical significance.

The results in *Paper 4* add to a research body of inconclusive findings regarding the heterogeneous associations between CRF and mental health for boys and girls: Ruggero et al. (2015), found baseline CRF to be inversely associated with follow-up depression levels among girls, not boys. Similarly, in the cross-sectional study by Greenleaf et al. (2010), higher levels of CRF indicated lower levels of depression among girls only. Similar findings were shown in the recent study by Janssen et al. (2020), in which there were stronger associations between CRF and internalizing problems among girls than among boys. Rieck et al. (2013), however, showed that boys, but not girls, with low CRF had higher odds of elevated depression than boys classified as having high CRF. Studies have also investigated whether the effect of PA interventions on mental health is moderated by sex. A meta-analysis found that the effect on mental health outcomes was larger for boys than girls in randomized studies; however, the effect was larger for girls than boys in non-randomized studies (Ahn & Fedewa, 2011). If the association between CRF and psychological difficulties is moderated by sex, the moderator may, in turn, be age-dependent, considering that adult women and men show similar favorable associations between CRF and mental health (Sui et al., 2009).

The mixed findings of the abovementioned studies, in conjunction with the present results limit the ability to make unequivocal conclusive arguments. Moreover, due to the scarcity of existing research examining the prospective associations between physical fitness and mental health outcomes, much of the following discussion relies on comparisons with studies measuring various forms of PA and exercise. Based on the subscale analyses that were carried out in *Paper 4*, potential explanations can be postulated: The inverse association between change in CRF and follow-up TDS was attributed to conduct problems, hyperactivity and peer problems. Firstly, conduct problems are more common among boys than girls (Button et al., 2007). Less conduct problems could be caused by improved self-regulation (Okado & Bierman, 2015), which implies a

behavioral mechanism (Lubans et al., 2016). Improved self-regulation could be caused by aerobic PA among adolescents (Laberge et al., 2012; Wills, Isasi, Mendoza, & Ainette, 2007). Additionally, the study by Stroth et al. (2009) found CRF to be associated with one of the subcategories of self-regulation, executive functions. Self-regulation is more common among girls than boys (Raffaelli, Crockett, & Shen, 2005) and therefore, the potential for change in self-regulation and, consequently, conduct problems, is larger among boys than girls. This lends support from the study by Lakes and Hoyt (2004), which also found reductions in conduct problems among boys only, and that the reduction was associated with an increase in self-regulatory skills, as the result of a martial arts program.

The inverse association between one-year change in CRF and mental health status among boys could also be attributed to lower hyperactivity. This may have occurred through a neurobiological mechanism (Lubans et al., 2016). Limited research elucidates the link between CRF and hyperactivity. However, CRF has been found to be beneficial for neurocognitive functioning (Moore, Drollette, Scudder, Bharij, & Hillman, 2014), which is a known impairment among individuals with ADHD, and especially among boys (Bálint et al., 2009). According to Gapin et al. (2011), there are many potential neurobiological ways that PA can influence hyperactivity, for instance by increasing blood flow to the frontal region of the brain, or by increasing the availability of dopamine and norepinephrine. This was supported in the recent review by Ng, Ho, Chan, Yong, and Yeo (2017), which concluded that "…moderately-to-intense aerobic exercise, is a beneficial and well-tolerated intervention for children and adolescents with ADHD". Unfortunately, the majority of studies investigating the effect PA may have on hyperactivity has only included a male population, which makes it difficult to evaluate potentially heterogeneous associations between boys and girls (Kamp, Sperlich, & Holmberg, 2014). However, that aerobic exercise may affect hyperactivity differently for boys and girls

has been observed previously, although explanations for the differences were unclear (Tantillo, Kesick, Hynd, & Dishman, 2002).

Lastly, the association between a one-year change in CRF and mental health status among boys could be attributed to peer problems. Similarly, in a previous study by Andersen et al. (2017), CRF was associated with social support and peers, although the effect size was interpreted to be trivial. Also, Lamb and Gulliford (2011), found lower levels of peer problems in children after an aerobic exercise program, and Sagatun, Søgaard, Bjertness, Selmer, and Heyerdahl (2007) found an inverse association between time spent in PA and peer problems among boys. The association between aerobic PA and peer problems may occur through a psychosocial mechanism, which specifically involves social interaction and relatedness (Lubans et al., 2016). Participation in team sports provides much opportunity for social interaction and the development of social skills (Eime, Young, Harvey, Charity, & Payne, 2013) and may be a potential explanation. Less peer problems may also be directly connected to reductions in hyperactivity and conduct problems, as adolescents who have problems in peer relationships commonly also display signs of hyperactivity (Bagwell, Molina, Pelham, & Hoza, 2001) or conduct problems (Woodward & Fergusson, 1999). The difference between boys and girls may be explained by previous results showing that boys perceive more social support, benefits, self-efficacy and fun from PA, compared to girls (Cardon et al., 2005). Thus, by experiencing more benefits from PA, boys are likely to be more active than girls, which in turn provides them with more perceived benefits, further prompting continued PA. This may lead boys into a self-perpetuating positive circle where PA and perceived benefits influence each other. Although this hypothesis is speculative and requires further exploration, it is supported by previous evidence indicating that adolescent girls experience less peer support than boys to participate in PA (Edwardson, Gorely, Pearson, & Atkin, 2013), and that adolescent

girls in general enjoy PE less than boys (Cairney et al., 2012; Laxdal & Giske, 2019).

The association between change in CRF and follow-up psychological difficulties among girls in the highest SES group was surprising. The p-value suggests very low compatibility with a null-hypothesis and the upper limit of the confidence interval suggests a concerning association: 100m increase in CRF was associated with up to 24% higher follow-up psychological difficulties than the mean levels in this subgroup (10.5). The result contradicts established knowledge: Aerobic PA, which is positive for CRF, is associated with improved mental health (Bailey et al., 2017). Furthermore, the result was attributed to hyperactivity and emotional problems. This is also incongruent with previous research showing that aerobic PA has a positive effect on hyperactivity and ADHD (Ng et al., 2017), and that meeting the recommended PA levels is associated with fewer emotional problems (Wiles et al., 2008). Importantly, these incongruences increase the likelihood of a spurious association. However, there may be an explanation: According to the annual report on Norwegian adolescents (Bakken, 2019), more girls than boys perceive stress and pressure, which is also more prevalent in high SES populations. Stress has been shown to be associated with hyperactivity (Biederman et al., 1995) and emotional problems (Moksnes, Moljord, Espnes, & Byrne, 2010). Therefore, it can be hypothesized that, within a female high SES-group, CRF may increase because of stress and pressure from high self-expectations to exercise more often than average, and to obtain body image ideals. Thus, the poor mental health status was not necessarily caused by an increase in CRF; instead, the variables may have mutually been affected by external stress and pressure, hence the association. Future studies that examine the association between physical fitness and mental health among adolescents should include SES as a covariate and potential moderator, to be able to substantiate, or contradict these findings.

7.1.2 *Muscular strength*

The results from *Paper 1* showed a cross-sectional association between muscular strength and TDS when the mixed model did not control for CRF. However, the association was not present when controlling for CRF. A possible explanation is that the participants with high CRF were also likely to have a relatively high muscular strength (Moliner-Urdiales et al., 2011) but the participants with high muscular strength did not necessarily have high CRF. A further interpretation is that muscular strength is not associated with psychological difficulties; they were only associated in the initial model because the fitness components are strongly correlated. *Paper 1* and other previous cross-sectional studies ruled out muscular strength as associated with mental health outcomes, when controlling for CRF (Andersen et al., 2017; Janssen et al., 2020; Kelly et al., 2010; Rieck et al., 2013; Yeatts et al., 2017). These studies, however, did not examine immigrant subgroups, which makes the findings of *Paper 4* important, as they provide evidence of an understudied subgroup.

The immigrant subgroup examined in *Paper 4* showed an inverse association between change in muscular strength and follow-up psychological difficulties. However, based on interpretations of the p-value and confidence interval, this is the least certain of the present findings. The p-value of the association ($p = .061$) is non-significant in the traditional sense, however, it still indicates that the data conforms more toward the association hypothesis, rather than the null hypothesis (Greenland et al., 2016). Furthermore, the confidence interval crosses the null value, but the limits indicate that the association with 1 SD muscular strength increase spanned from 38% lower- to 0.9% higher follow-up TDS than the immigrant means (10.7). The upper limit represents an association of small clinical significance, while the lower limit represents a potentially relevant and clinically important association (Goodman & Goodman, 2009). Given the scarcity of similar previous research, it can only be speculated why the present results indicated that increased

muscular strength may be favorable for conduct problems and hyperactivity, among immigrant adolescents specifically. A possible explanation is a psychosocial mechanism (Lubans et al., 2016) revolving around "the self" a collective term that includes related concepts such as self-esteem, self-perception and self-efficacy (Collins et al., 2019). Studies have found muscular strength to be associated with self-esteem and similar concepts such as self-perception, perceived physical appearance and physical self-worth (Lubans & Cliff, 2011; Smith et al., 2014). These findings have also been supported by recent research suggesting a strong link between strength training and "the self" in adolescents (Collins et al., 2019). In turn, lower self-esteem has been associated with both conduct problems (Ha, Petersen, & Sharp, 2008) and hyperactivity (Edbom, Lichtenstein, Granlund, & Larsson, 2006). The reason why the present association was found only among immigrant adolescents could be due to immigrants having lower self-esteem than their non-immigrant peers (Bankston & Zhou, 2002). To discuss potential reasons why immigrant adolescents may have poorer sense of self than their non-immigrant peers is beyond the scope of this thesis. However, when considering the challenging process of adolescent development (Christie & Viner, 2005; Steinberg & Morris, 2001), in combination with the challenges to peer relationships and identity development that immigrated adolescents experience (Oppedal, Røysamb, & Heyerdahl, 2005; Virta, Sam, & Westin, 2004), it appears that there are greater threats to immigrants' sense of self than their native peers. Research suggests that how adolescents perceive their physical appearance is the most important aspect of their overall sense of self (Harter, 2003; Harter & Whitesell, 2001). Furthermore, positive peer relationships and belonging to a peer group are associated with adolescent identity development (Ragelienė, 2016). However, perceived physical appearance, feelings of belonging to a peer group and identity development can be distorted by discrimination (Virta et al., 2004). Thus, the present findings in concert with previous research, suggest the possibility that improving muscular strength can attenuate this distortion

and prevent negative senses of the self. The societal importance of integration and acculturation of immigrants warrants more research to corroborate the findings.

7.1.3 *Body composition*

The present results showed no cross-sectional or prospective association between BMI and psychological difficulties. This is in congruence with similar studies that control for CRF in the analyses, and do not find associations between body composition and depression, or similar mental health outcomes (Janssen et al., 2020; Kelly et al., 2010; Rieck et al., 2013; Ruggero et al., 2015; Yeatts et al., 2017). There are examples of studies with adolescent populations that find body composition to be associated with appearance-related mental health outcomes, such as self-esteem, physical self-worth, body image and body satisfaction (Greenleaf et al., 2010; Lubans & Cliff, 2011; Voelker, Reel, & Greenleaf, 2015). It might be a psychosocial mechanism, through which body composition influences appearance-related mental health outcomes, but not depression. According to Markey (2010), body image is not static; it refers to "…how individuals feel about themselves as they inevitably undergo physical changes across the lifespan". Considering the major physical changes that characterize adolescence, it is a period in which body image can be vulnerable. Body image is also an important aspect of adolescent development, considering that perception of physical appearance may constitute the most important aspect of adolescents' sense of self (Harter, 2003; Harter & Whitesell, 2001). Body image and -satisfaction/dissatisfaction is largely influenced by cultural appearance ideals, peers, parents and social comparison (Grabe et al., 2008; Lawler & Nixon, 2011; Schaefer & Salafia, 2014). It is therefore likely that, among adolescents especially, appearance-related components of physical fitness, such as body composition, are associated with body image and similar concepts. Regrettably, the instruments used in the School in Motion study do not measure these subsets of mental

health and it is therefore assumed with caution that the association might exist within the present population. However, the potential link between physical fitness and appearance-related mental health outcomes is important to elucidate, considering the link between body image and depression, which has been shown to persist from adolescence to adulthood (Blashill & Wilhelm, 2014). Equally important is the elucidation of a major limitation among several studies of this research area: The studies that found associations between body composition and mental health outcomes, did not control for CRF in their analyses (Greenleaf et al., 2010; Lubans & Cliff, 2011). This limitation is also prevalent in similar studies of adult populations: Close examination of two systematic reviews that found body composition to be associated with depression, revealed that none of the reviewed studies controlled for CRF (Luppino et al., 2010; Pereira-Miranda, Costa, Queiroz, Pereira-Santos, & Santana, 2017). Not controlling for the potentially confounding effect of CRF, or muscular strength, can potentially lead to erroneous conclusions. Thus, future studies should control for all three components of physical fitness, or at the very least, researchers should be mindful of the limiting consequences of only measuring body composition, when assessing associations with mental health outcomes.

In summary, the nature of the association between physical fitness components and mental health among students in Norwegian lower secondary school, is complex. The findings suggest that the associations depend on moderators such as sex and immigrant status. Moreover, the findings also suggest that the moderators influence different components of physical fitness (e.g., sex and SES → CRF, immigrant status → muscular strength).

7.2 Research question 2

How is a complex school-based PA intervention implemented in Norwegian 9th grades and what influences the implementation process?

This question is elucidated in *Paper* 2, which aimed to evaluate the process of how intervention model 2, Don't worry, be happy, was implemented.

The main findings show large differences between schools regarding how DWBH was implemented and how various factors influenced the implementation. Schools 1 and 4 made minor adaptations in the way DWBH was organized, and these were positively received by the students. Intervention characteristics, spacious facilities, scheduling and participant- and provider characteristics positively influenced all aspects of implementation. School 2 made major adaptations to how DWBH was scheduled, which reduced both responsiveness and fidelity. Additionally, limited facilities, participant- and provider characteristics negatively impacted fidelity, quality and dose received. School 2 was the only school where the intervention was negatively received. School 3 made one major adaptation in how DWBH was organized, and it was poorly received by the students. The intervention itself was otherwise positively received. Limited facilities and scheduling negatively impacted fidelity. Intervention-, participant- and provider characteristics positively influenced responsiveness, quality and perhaps also dose received.

In Schools 1, 3 and 4, the teachers decided to have two identical BH lessons rather than DW and BH. The providing teachers decided on this adaptation because they thought it fit with the purpose of the intervention and because they had the opportunity to do it. Students and teachers agreed that the adaptation was beneficial. In School 2, however, the adaptations were made primarily from contextual limitations: All grades and classes followed a fixed schedule and they could not reorganize only for 9th grade students. Therefore, two intervention lessons were added to the schedule instead of only one, which reduced students' leisure time and caused a negative response before the intervention had even started. Moreover, School 2 had inadequate facilities relative to their number of students. Therefore, most of the students had to do their BH activities off

school grounds. These factors, in combination with limited teacher supervision, likely led to frequent truancy among students in School 2. There were also contextual limitations that caused adaptations in School 3: There were inadequate facilities relative to the number of students, which compelled the supervising teacher to carry out BH in separate classes. However, because of poor scheduling, the gymnasium was occupied during one of the intervention lessons, which prevented one class from using the gymnasium in their BH lessons. The adaptation was not positively received among students. These adaptations and their respective reactions can be elucidated by Moore, Bumbarger, and Cooper (2013) who claim that adaptations can be either positive, neutral or negative, and be carried out for either logistical or philosophical reasons. An intervention can be adapted to fit the context and positively influence implementation (Berkel et al., 2011; Durlak & Dupre, 2008), which is what happened in Schools 1 and 4, and to some extent in School 3, where they made positive adaptations for philosophical reasons. In School 2 and partly in School 3, however, negative adaptations for logistical reasons negatively impacted implementation. The findings involving adaptations indicate that schools that were likely (because of their preconditions) to succeed in implementing DWBH anyway, made the positive adaptations. Conversely, the schools that were less likely (because of their preconditions) to succeed in implementing DWBH made the negative adaptations.

Facilities were an important frame factor that differentiated the schools and influenced how DWBH, was adapted differently in each school. In combination with the facilities, the implementation process continued to develop from the dynamic interaction between intervention components, teachers and students. Furthermore, this development occurred differently on multiple levels: the school level (e.g. different facilities, scheduling and teachers), class level (e.g. the scheduling problem in School 3, where one of the participating classes did not have facilities during DW), activity-group level (e.g. few groups within a school

pursued the same activities) and student level (e.g. students lost interest and changed groups). This development fits the concept of emergence, described in complexity theory as unpredictable behavior that comes from complexity, irrespectively of previous knowledge (Mason, 2008). Moore et al. (2019) have also argued that introducing a complex intervention in a complex system poses an almost infinite number of uncertainties. Therefore, and despite plenty of qualitative information, conclusive remarks cannot be stated regarding how suitable DWBH is for all lower secondary schools in Norway. Moreover, addressing whether DWBH "worked" is not within the scope of *Paper 2*. The process evaluation, however, provides additional information about whom, under what circumstances and why (Bonell et al., 2012; Deaton & Cartwright, 2018), to improve the ability to interpret the outcome results (Oakley et al., 2006).

To summarize how DWBH was implemented and what influenced implementation process; the degree of implementation varied greatly and depended upon several factors, with facilities emerging as potentially the most important factor. Moreover, the results indicate that the majority of students and teachers found DWBH to be relevant and enjoyable when 1) DWBH had few or no consequences for students' leisure time; 2) DWBH was executed with adequate facilities; 3) DWBH was provided by teachers who were present and cared about what their students did; and 4) adaptations were perceived positively and did not negatively impact on 1), 2) or 3).

7.3 Research question 3

To what extent can two school-based PA interventions affect the mental health of Norwegian 9th graders?

Paper 3 attempts to answer this question, by presenting the intervention outcomes of the School in Motion study. The aim of *Paper 3* was to assess the effect of two school-based PA interventions on adolescents'

mental health. The complete case results indicated that psychological difficulties for the overall population were not affected by the interventions. Subgroup analyses, however, showed beneficial effects from both interventions: Intervention model 1 reduced psychological difficulties in the subgroup with the highest baseline levels of TDS, and both interventions prevented psychological difficulties from increasing in the immigrant subgroup. Analyses of the SDQ subscales also revealed an opposite effect: Intervention model 2 caused peer problems to increase among non-immigrant girls and in the subgroup with borderline TDS at baseline.

Although the immigrant subgroups' mixed model p-values were non-significant in the traditional sense, they were low enough to indicate that the data conformed more to the hypothesis of an effect, than the null hypothesis (Greenland et al., 2016). Furthermore, compared to the control group, the difference in change of 95% of intervention group 1 spanned from a 3.5 points lower increase to a 0.3 points higher increase in TDS. The equivalent difference in change for 95% of intervention group 2 spanned from a 4.4 points lower increase to a 0.2 points higher increase. The two upper limits of 0.3 and 0.2 points, respectively, suggest an increase of no clinical importance. The lower limits, however, represent effects of potentially substantial clinical significance (Goodman & Goodman, 2009). Therefore, despite p-values above the traditional significance level and confidence intervals containing null-values; the interventions may have assisted in preventing a clinically significant risk-increase for developing a mental disorder for a majority of the immigrant subgroup.

Respectively, 85% and 83% in intervention groups 1 and 2 had normal levels of psychological difficulties. The absence of an overall effect is therefore not surprising, and it is possible that a ceiling effect has occurred. Ceiling effects were also suspected in the studies by Eather et al. (2016), Smith et al. (2018) and Christiansen et al. (2018), which compare closely to *Paper 3*. One possible explanation is the intervention

period: The School in Motion interventions lasted 29 weeks, longer than in similar studies (Eather et al., 2016; Lubans et al., 2016; Smith et al., 2018). However, it might have been too short to attenuate an overall increase in psychological difficulties among generally healthy adolescents. Furthermore, although 80-81% of the dose was registered as delivered, analyses on PA outcomes show that intervention model 1, "Active Learning", slightly increased school-time PA levels, while DWBH did not (Kolle et al., 2020, submitted). This difference may have been caused by the design of DWBH: All three components of Active Learning were teacher-led and were anticipated to be performed with moderate to high intensity. DWBH, however, contained two lessons that were mainly student-led. The process evaluation in *Paper 2* substantiates the hypothesis that the extensive freedom and student-led activities that characterized DWBH sometimes led to truancy. Additionally, the intervention specified that students were allowed to choose their preferred activity, which involved everything from low intensity walking to high intensity soccer. These factors may have resulted in a dichotomization among the DWBH participants, characterized by physical activity and inactivity. The absence of an overall effect may also be partially explained by the control population, who also were physically active: Lower secondary school students in Norway attend at least two PE lessons per week and can opt in for the elective subject of physical activity and health (often organized weekly in 90-minute lessons). In addition, 63% of lower secondary school students in Norway participate in leisure time sports (Bakken, 2019). Among upper secondary school students, however, only 40% participate in leisure time sports and it is possible that an intervention continuing into this period of adolescence would have shown an effect in the overall population.

7.3.1 *Abnormal psychological difficulties at baseline*

TDS decreased in all abnormal TDS at baseline subgroups respective to intervention group 1, 2 and control, indicating a regression to the mean.

However, the abnormal TDS at baseline subgroup within intervention group 1 displayed a reduction in TDS that was more than twice as big as the reduction in the corresponding control subgroup. The subgroup's mean reduction from Active Learning was nearly 3 points larger than the reduction in the corresponding control subgroup, which may represent a clinically important difference (Goodman & Goodman, 2009). The results concur with previous studies that also found effects in subgroups with poor mental health at baseline (Christiansen et al., 2018; Eather et al., 2016; Smith et al., 2018).

The effect on TDS could be attributed to reductions in conduct problems and hyperactivity. A possible explanation for the reduction in conduct problems is the behavioral mechanism (Lubans et al., 2016), which also was proposed as an explanation for part of the findings in *Paper 4*. To reiterate, aerobic PA can positively influence self-regulation (Laberge et al., 2012; Wills et al., 2007) and coping (Wijndaele et al., 2007), which may act as a preventive factor against conduct problems (Ebata & Moos, 1991; Okado & Bierman, 2015). There is conflicting evidence against this finding also, for example the meta-analysis by Ahn and Fedewa (2011), which did not show an association between PA and conduct problems. The reduction in hyperactivity can be explained by the neurobiological mechanism (Lubans et al., 2016), another possible explanation for findings in *Paper 4*. According to the review by Gapin et al. (2011), PA can be beneficial for hyperactivity/ADHD by increasing cerebral blood flow, availability of dopamine and norepinephrine, and brain derived neurotrophic factor.

7.3.2 *Immigrants*

The immigrant subgroups that received Active Learning and DWBH displayed a smaller increase and no increase in TDS, respectively, compared to the corresponding control subgroup. Mainly, the effect from Active Learning could be attributed to reductions in emotional problems, and the effect from DWBH could be attributed to reductions in emotional

problems and a prevented increase in hyperactivity. Possibly, the psychosocial mechanism, social support, may explain the results from both interventions (Lubans et al., 2016): Social support has been associated with sports participation (Babiss & Gangwisch, 2009), which can explain the results from Active Learning. Moreover, social support may have been improved from DWBH, considering the interventions was designed to facilitate positive social relationships. Immigrant adolescents experience less social support than non-immigrants, and therefore might have a larger potential for change (Oppedal & Røysamb, 2004). Social support has been inversely associated with hyperactivity (Mastoras, Saklofske, Schwean, & Climie, 2015) and can also stave off emotional problems (Garnefski & Diekstra, 1996), thus potentially explaining the results from both interventions.

In the discussion of *Paper 4* (Chapter 7.1.2), the sense of self was suggested as a potential explanatory mechanism for the prospective association between muscular strength and psychological difficulties among immigrant adolescents. This might also help explain the interventions' effects. To reiterate, adolescent immigrants may have lower self-esteem than their non-immigrant peers (Bankston & Zhou, 2002) and may therefore have more potential for change. Furthermore, PA is associated with adolescent self-esteem (Dale, Vanderloo, Moore, & Faulkner, 2019) and it is possible that immigrants' sense of self improved as a result of the positive social context of DWBH. In turn, self-esteem is associated with hyperactivity (Edbom et al., 2006) and emotional health (Moksnes & Espnes, 2012). Identity development is an important part of adolescence and can be influenced by positive peer relationships, self-esteem, self-perception and self-efficacy (American Psychology Association, 2002; Collins et al., 2019; Ragelienė, 2016). However, racial discrimination against immigrants occurs and can influence these and other connected outcomes (Virta, Sam, & Westin, 2004). Although more data is needed to support these speculations, it is

possible that both interventions contributed to reduce negative influences on immigrants' sense of self.

Immigrant adolescents are an understudied population when it comes to investigating the effects of PA on mental health. However, there exist somewhat comparable studies: For instance, a study on an adult immigrant population in Sweden investigated the effect of a four-month lifestyle intervention. The intervention positively influenced mental health, and the authors emphasized increased PA and social support as potential causes (Siddiqui, Lindblad, Nilsson, & Bennet, 2019). Furthermore, in a qualitative study, "enhanced self-confidence, happiness, and lower stress" were frequently reported as experienced benefits from PA among adult and adolescent immigrants of different origins in USA (Wieland et al., 2015).

7.3.3 *Detrimental effects*

In the subgroups non-immigrant girls and borderline TDS at baseline that received DWBH, peer problems increased, while their respective control counterparts displayed decreases. The differences in change were small to moderate but gave reason for concern. Scoring high on peer problems is characterized by the SDQ as being solitary, having few friends, being bullied by others and getting along better with adults than with peers. The negative effects were surprising, considering that the intervention was designed to facilitate social relationships through PA. Additionally, participation in team sports has been associated with fewer mental problems than participation in individual sports (Breistøl, Clench-Aas, Van Roy, & Raanaas, 2017), and a recent review by Pels and Kleinert (2016) concluded that PA could contribute to reducing loneliness. The process evaluation of DWBH, outlined in *Paper 2*, offers a possible explanation for the negative effect among girls: The formation of activity groups could lead to some girls feeling ostracized, while planning and cooperating within the group sometimes led to disagreements, which could cause one or several girls to leave their groups. Previous research

on sex differences in adolescent peer relationships has shown that social anxiety, expressed by a "fear of negative evaluation from peers, and more social avoidance and distress in new situations" (La Greca & Lopez, 1998), is more prevalent in girls than in boys. Furthermore, while boys tend to thrive in social groups, girls tend to place more emphasis on dyadic relationships (Prinstein, Borelli, Cheah, Simon, & Aikins, 2005), perhaps because they experience less conflicts in these relationships. According to Xie, Swift, Cairns, and Cairns (2002), conflicts among girls involving social aggression, e. g., exclusion, isolation and gossiping, most often occur in groups of four or more members, a common group size in the DWBH intervention. The potential sources of conflict for the non-immigrant girls may also have been the reason for the negative effect in the borderline TDS subgroup; however, it is unclear why the negative effect occurred specifically in this subgroup.

7.4 Methodological considerations

This thesis contains many strengthening characteristics. The School in Motion project recruited a large study population; quantitative and qualitative methodological approaches have been used to reach the thesis' aims; three separate components of physical fitness were objectively measured; a widely used and validated instrument measured mental health; and analyses included several important covariates and subgroups. However, there are also several limitations that must be considered. The following sections present the methodological limitations, their circumstances and the considerations that were made, in terms of how the limitations may have influenced the research process.

7.4.1 *Quantitative studies*

The main limitations of *Papers 1, 3 and 4,* is the large number of missing values. Out of 2084 consented participants from all schools participating in School in Motion, only 925 participants completed both baseline and follow-up testing. The missing values were primarily handled by

complete case analyses and supplemented with multiple imputation. Complete case analysis is debated because it excludes much observed information (Bartlett, Carpenter, Tilling, & Vansteelandt, 2014). According to Hughes, Heron, Sterne, and Tilling (2019), complete case analysis can potentially produce unbiased results if data are missing at random, missing completely at random and not missing at random. Others, however, argue that complete case analyses are valid only when data are missing completely at random, and in case of other missingness mechanisms, multiple imputation is a superior alternative (Pedersen et al., 2017; Sterne et al., 2009). On the other hand, when large amounts of data are imputed, which was the case in the present analyses, multiple imputation can also produce biased results (Hughes et al., 2019; Lee & Carlin, 2012). Moreover, multiple imputation can create another problem if used on cluster randomized groups and subgroups, by skewing the imputed values toward the mean (Sullivan, White, Salter, Ryan, & Lee, 2018) and thus attenuate potential effects and associations. In reality, there are no satisfactory solutions when data are missing (Altman & Bland, 2007), thus missing data handling is more about choosing the lesser evil and transparent presentation. Ideally, the amount of missing data should have been prevented when testing; however, this was proven difficult in the present data collections. Because testing was voluntary, no participant was forced to complete the tests, which caused some participants to opt out. To give an anecdotal example, there were many students who stated that they did not want to run the CRF test and many did not want to measure their weight. Moreover, the School in Motion study measured many other outcome variables via questionnaire. Thus, SDQ was one of many components in a large and extensive questionnaire, which took a long time to complete (45-60 minutes). The missing data from the SDQ may be caused by participants who did not bother to complete or did not bother to fill out all questions. Future studies that use questionnaires composed of several instruments, should consider that the size of the questionnaire may be detrimental for

participation. Another consideration is the possibility of offering an incentive, to motivate participants to complete the questionnaire.

The main findings of *Paper 3 and 4* are based on subgroup analyses and there is a risk that these findings are spurious. On one hand, it is of this author's opinion that by adhering to the critical criteria set by Sun et al. (2012), the credibility of the subgroup findings remains satisfactory. On the other hand, considering the small size of the small subgroups and the potential influence of missing values, the subgroup results are uncertain and need confirmation in future studies.

As explained in Chapter 5.1.3 Outcome measures, the running field distance in the CRF-test had to be reduced from 20 m to 16 m, due to spatial limitations in many gymnasiums. This reduction is a limiting factor as it makes our CRF-results incomparable to results from other studies that use the Andersen-test. However, this limitation did not affect the main results since running distance in meters was used as the measurement unit, and not estimated maximum oxygen uptake.

There are some concerns regarding the use of the self-report version of the SDQ. Respondents who are required to evaluate themselves positively or negatively via self-report questionnaires have been shown to answer in a socially desirable way (Leising, Locke, Kurzius, & Zimmermann, 2016). Another concern is the potentially low sensitivity to change, explained by Goodman et al. (2000): SDQ completed by parent, teacher and the child itself has the highest sensitivity. Moreover, the recommendation is that if one of these measures has to be dropped, it should be the self-report measure. This constitutes a major uncertainty for the results, considering that SDQ is the main outcome variable. The key reason for using self-report was to avoid placing an unnecessary load on the teachers, who already were required to provide the interventions. To have parents, students and teachers complete the SDQ was not considered as a possible solution. The internal consistency results pertaining to SDQ were quite low and are similar to Italian (Riso et al.,

2010), Finnish (Koskelainen, Sourander, & Vauras, 2001) and Dutch (Muris et al., 2004) study results. Internal consistency results tend to favor native English speakers (Goodman et al., 2003; Ruchkin, Jones, Vermeiren, & Schwab-Stone, 2008), which suggests that they understand the SDQ statements better than populations who use translated questionnaires. Age is also a factor, as the internal consistency is lower for younger adolescents, such as the present population, compared to older adolescents as examined in studies by Bøe, Hysing, Skogen, and Breivik (2016) and Sagatun et al. (2007).

The cross-sectional and prospective cohort designs of *Papers 1 and 4* limit the ability to make causal inferences, while the RCT design of *Paper 3* permits causal inferences. However, the use of RCTs to evaluate the effect of a complex public health intervention is debated (Byrne, 2013; Hawe et al., 2004) and to quote Bonell et al. (2012), "the magnitude of effects in complex public health interventions can be much more modest, gradual, and entangled with other factors". While randomization is assumed to eliminate systematic differences between units, the process evaluation of *Paper 2,* indicated that at least three variables (frame factors, participant characteristics and provider characteristics) varied greatly between schools. The variation may have caused systematic differences between clusters within the intervention group and the differences likely influenced DWBH outcomes (Durlak & Dupre, 2008). The differences caused DWBH to be implemented differently, to the extent that students in different schools received different interventions. It is possible that if the pre-determined inclusion criteria also included adequate facilities, only schools that could accommodate the intervention would have been invited. This would have reduced the representativeness of the schools, but on the other hand, the participants would be more likely to have received the intervention as intended, and without much variation. The intervention period of 29 weeks may also have been too short. Successful implementation of interventions in schools can take between three and five years (Rimm-

Kaufman, Fan, Chiu, & You, 2007; Sugai & Horner, 2006). Moreover, even if the implementation was adequate the entire period, 29 weeks might not be enough time to attenuate an overall increase in psychological difficulties among generally healthy adolescents.

7.4.2 *Qualitative study*

Ideally, data collection should have occurred more than once (Moore et al., 2015). Limited time and resources only allowed collection at one time point and restricted the number of schools that could be qualitatively evaluated. Implementation can change over time (Dusenbury et al., 2003), therefore it is important to note that the present findings might not represent implementation throughout the intervention period. Moreover, considering that the qualitative data collection spanned over two months, the implementation process in the first school evaluated might have been at a different stage than in the last school evaluated. Possibly, a qualitative data collection at the end of the intervention period would reflect a more successful implementation. Unfortunately, this was not possible because the quantitative data collection was scheduled at the end of the intervention period, thus, there would not be sufficient time to do both simultaneously.

7.5 Implications

Before the implications of this thesis can be outlined, it is important to point out that this thesis does not intend to recommend additional school-time PA as a miracle cure against all mental health problems, academic struggles and hardships related to adolescent development. There will always exist specific cases that should be handled individually, with carefully thought out approaches that go beyond the topics discussed in this thesis.

The results indicate that school-time PA programs are beneficial for the mental health of adolescent immigrants and adolescents with high levels

of psychological difficulties. Moreover, increasing muscular strength may be beneficial for psychological difficulties among adolescent immigrants, and increasing CRF may be beneficial for psychological difficulties among adolescent boys. The Active Learning intervention, or elements of it, could be implemented as a strategy to positively influence the mental health of immigrants and students who are at risk of developing mental problems. Furthermore, it is possible that on a longer term, the PA program may also show benefits for the overall population, considering that mental problems increase and PA decrease throughout adolescence (Bakken, 2019; Dalene et al., 2018). Special notice should be taken of the results that pertain to the immigrant subgroup. The hypothesized mechanisms used to explain the results suggest that the immigrants' experience of adolescence is characterized by less positive peer relationships and threats to the sense of self and identity development. If school-time PA improves on this experience, it should be strongly considered by school stakeholders. The DWBH intervention especially, showed promising effects in the immigrant subgroup. As a consequence, aspects of the DWBH intervention, such as forming small activity groups to facilitate positive experiences through PA could be implemented in school as a program of its own, or as part of ordinary PE lessons. The design and execution of DWBH should also be thoroughly revised and evaluated in other studies, to avoid the negative effects on peer problems that were displayed among non-immigrant girls and those with borderline psychological difficulties. It is possible that more teacher supervision can prevent conflicts from occurring, but this can only be speculated.

Regarding the feasibility of the evaluated PA programs, the results are generally favorable. However, it is important to note that schools received extra funding to implement one additional hour to their weekly schedules. It is therefore unlikely that the programs, as they are, can be implemented without this funding. Additionally, programs such as DWBH may be too demanding for some schools that have inadequate

facilities, relative to the number of students. Attempts to implement DWBH in schools that are not equipped for having entire grades in simultaneous activity can have detrimental consequences for how participants view the program and how it is implemented. However, it is feasible to implement certain aspects of the DWBH intervention as part of ordinary PE lessons, as mentioned in the previous paragraph. The physically active academic lessons of the Active Learning intervention are likely the most feasible aspect of the two interventions, as they can be incorporated into other subjects. However, this also requires resources, as teachers need to learn how to incorporate PA in each theoretical subject. Both interventions required the schools to allocate time from other subjects to make room for PA. This can also be considered feasible, although at least one school failed to comply with the requirement.

The implications that may come from this thesis should be considered in light of the educational aims and responsibilities, curricular development and political initiatives that are outlined in Chapter 2. Firstly, these results should have implications for how the new core curriculum in Norway, and its interdisciplinary topic "Health and life skills" (The Norwegian Directorate for Education and Training, n. d.) is implemented. Secondly, implementing the PA programs, or aspects of them, would also assist schools in fulfilling their responsibility of contributing to students' health and well-being ("The Education Act," 1998a). Thirdly, by implementing the PA programs, or part of them, schools can fulfill the objectives in "The Public Health Message", which stated that the government had a responsibility to protect children and adolescents from the dangers of inactivity (Ministry of Health and Care Services, 2015).

7.6 Recommendations for future research

This thesis' literature review, results, discussion and practical implications all form a base from which future research should be directed.

The first recommendation revolves around the novel findings regarding how adolescent immigrants especially can benefit from school-time PA and, possibly, increased muscular strength. Discussing the role of school in issues related to immigration is beyond the scope of this thesis. However, schools are important arenas for the integration and acculturation of immigrated children and adolescents. Much effort should be put into researching whether school-time PA is an approach that can benefit the mental health of this population, or if the present findings were spurious.

The second recommendation is to re-evaluate the DWBH intervention, with the purpose of avoiding negative effects for any subgroup.

The third recommendation is to further explore the relationship between physical fitness and mental health among adolescents, using longitudinal designs. Research is scarce, and improved knowledge of this relationship can be important for how schools, or other societal institutions, work to promote adolescent health.

Other, more general, but equally important recommendations for the research field include:

- Always perform a process evaluation of a complex RCT and connect the findings to the intervention outcomes.
- If possible, investigate potential moderators of the relationships between PA, physical fitness and mental health.
- Future examinations of the relationship between physical fitness and mental health should include more than one measure of physical fitness. They should also include more than one measure

of mental health and, if possible, measure hypothesized mechanisms to examine whether they mediate the relationship between PA, physical fitness and mental health.

8 Conclusion

The aims of this thesis were to elucidate various aspects of the relationship between school-based PA interventions, physical fitness and mental health among adolescents, and the feasibility of school-based PA interventions in lower secondary school. These topics have been elucidated through four research papers, using different methodological approaches.

CRF was associated with mental health and improved CRF over one year was associated with mental health among boys. In contrast, improved CRF over one year was associated with poorer mental health among girls with high socioeconomic status. Although this might be explained by the stress and pressure experienced by this demographic, the findings nuance the established knowledge of the association between PA and mental health and exemplify a context in which other measures than PA should be considered to promote mental health. Improved muscular strength over one year was associated with mental health among immigrants, which constitutes a novel finding in an understudied subgroup. Implementing complex school-based PA programs in lower secondary school is feasible, with certain very important caveats, depending on the intervention characteristics, the school's frame factors, provider characteristics and student characteristics. There was a favorable causal relationship between the interventions and mental health among immigrants and those with high levels of psychological difficulties at baseline. In contrast, findings also show a detrimental causal relationship between one of the interventions and peer relationships among girls and those with moderate levels of psychological difficulties at baseline. This negative effect was likely caused by specific aspects of the intervention, unrelated to PA per se. The collective findings of this thesis indicate that improving physical fitness and implementing school-based PA programs can be beneficial for mental health promotion among Norwegian adolescents.

9 References

Aadland, E., Andersen, L. B., Lerum, Ø., & Resaland, G. K. (2018). The Andersen aerobic fitness test: New peak oxygen consumption prediction equations in 10 and 16-year olds. *Scandinavian Journal of Medicine & Science in Sports*, 28(3), 862-872. doi: 10.1111/sms.12985

Abebe, D. S., Lien, L., & Hjelde, K. H. (2014). What We Know and Don't Know About Mental Health Problems Among Immigrants in Norway. *Journal of Immigrant Minority Health, 16*(1), 60-67. doi: 10.1007/s10903-012-9745-9

Agans, J. P., Sävfenbom, R., Davis, J. L., Bowers, E. P., & Lerner, R. M. (2013). Positive Movement Experiences: Approaching the Study of Athletic Participation, Exercise, and Leisure Activity through Relational Developmental Systems Theory and the Concept of Embodiment. *Advances in Child Development and Behavior, 45*, 261-286. doi: 10.1016/B978-0-12-397946-9.00010-5

Ahn, S., & Fedewa, A. L. (2011). A Meta-analysis of the Relationship Between Children's Physical Activity and Mental Health. *Journal of Pediatric Psychology, 36*, 385-397. doi: 10.1093/jpepsy/jsq107

Ainsworth, B. E., Haskell, W. L., Whitt, M. C., Irwin, M. L., Swartz, A. M., Strath, S. J., . . . Emplaincourt, P. O. (2000). Compendium of physical activities: an update of activity codes and MET intensities. *Medicine and Science in Sports and Exercise, 32*(9 SUPP/1), S498-S504. Doi: 10.1097/00005768-200009001-00009

Allmark, P., Boote, J., Chambers, E., Clarke, A., McDonnell, A., Thompson, A., & Tod, A. M. (2009). Ethical Issues in the Use of In-Depth Interviews: Literature Review and Discussion. *Research Ethics, 5*(2), 48-54. doi: 10.1177/174701610900500203

Altman, D. G., & Bland, J. M. (2007). Missing data. *BMJ, 334*(7590), 424. doi:10.1136/bmj.38977.682025.2C

Amrhein, V., Greenland, S., & McShane, B. (2019). Scientists rise up against statistical significance. *Nature, 567*, 305-307. doi: 10.1038/d41586-019-00857-9

Andersen, J. R., Natvig, G. K., Aadland, E., Moe, V. F., Kolotkin, R. L., Anderssen, S. A., & Resaland, G. K. (2017). Associations between health-related quality of life, cardiorespiratory fitness, muscle strength, physical activity and waist circumference in 10-year-old children: the ASK study. *Quality of Life Research, 26*(12), 3421-3428. doi: 10.1007/s11136-017-1634-1

Andersen, L., Andersen, T. E., Andersen, E., & Anderssen, S. (2008). An intermittent running test to estimate maximal oxygen uptake: the Andersen test. *Journal of Sports Medicine and Physical Fitness, 48*(4), 434-437.

American Psychiatric Association, (2013). *Diagnostic and statistical manual of mental disorders (DSM-5®)*: American Psychiatric Pub. Retrieved from https://cdn.website-editor.net/30f11123991548a0af708722d458e476/files/uploaded/DSM%2520V.pdf

American Psychological Association. (2002). *Developing Adolescents: A Reference for Professionals*. Washington, DC.

American Psychological Association (n. d.). Psychosocial. Retrieved from https://dictionary.apa.org/psychosocial

Arseneault, L., Bowes, L., & Shakoor, S. (2010). Bullying victimization in youths and mental health problems: 'much ado about nothing'? *Psychological Medicine, 40*(5), 717-729. doi: http://dx.doi.org/10.1017/S0033291709991383

Babiss, L. A., & Gangwisch, J. E. (2009). Sports participation as a protective factor against depression and suicidal ideation in adolescents as mediated by self-esteem and social support. *Journal of Developmental & Behavioral Pediatrics, 30*(5), 376-384. doi: 10.1097/DBP.0b013e3181b33659

Bagwell, C. L., Molina, B. S. G., Pelham, W. E., & Hoza, B. (2001). Attention-Deficit Hyperactivity Disorder and Problems in Peer Relations: Predictions From Childhood to Adolescence. *Journal of the American Academy of Child & Adolescent Psychiatry, 40*(11), 1285-1292. doi: https://doi.org/10.1097/00004583-200111000-00008

Bailey, A. P., Hetrick, S. E., Rosenbaum, S., Purcell, R., & Parker, A. G. (2017). Treating depression with physical activity in adolescents and young adults: a systematic review and meta-analysis of randomised controlled trials. *Psychological Medicine, 48*(7), 1-20. doi: 10.1017/S0033291717002653

Bakken, A. (2017). *Ungdata. Nasjonale resultater 2017*. Retrieved from Oslo: http://www.hioa.no/Om-HiOA/Senter-for-velferds-og-arbeidslivsforskning/NOVA/Publikasjonar/Rapporter/2017/Ungdata-2017

Bakken, A. (2019). *Ungdata. Nasjonale resultater 2019*. Retrieved from Oslo: https://fagarkivet.oslomet.no/en/item/ungdata-2019-nasjonale-resultater

Bálint, S., Czobor, P., Komlósi, S., Mészáros, Á., Simon, V., & Bitter, I. (2009). Attention deficit hyperactivity disorder (ADHD): gender- and age-related differences in neurocognition. *Psychological Medicine, 39*(8), 1337-1345. doi: 10.1017/S0033291708004236

Bandura, A. (1982). Self- efficacy mechanism in human agency. *American Psychologist, 37*(2), 122-147. doi: 10.1037/0003-066X.37.2.122

Bankston, C. L., III, & Zhou, M. (2002). Being well vs. doing well: Self-esteem and school performance among immigrant and nonimmigrant racial and ethnic groups. *The International Migration Review, 36*(2), 389-415. doi: http://dx.doi.org/10.1111/j.1747-7379.2002.tb00086.x

Barriball, K. L., & While, A. (1994). Collecting data using a semi-structured interview: a discussion paper. *Journal of Advanced Nursing, 19*(2), 328-335. doi: 10.1111/j.1365-2648.1994.tb01088.x

Bartlett, J. W., Carpenter, J. R., Tilling, K., & Vansteelandt, S. (2014). Improving upon the efficiency of complete case analysis when covariates are MNAR. *Biostatistics, 15*(4), 719-730. doi: 10.1093/biostatistics/kxu023

Becker, D. R., McClelland, M. M., Loprinzi, P., & Trost, S. G. (2014). Physical Activity, Self-Regulation, and Early Academic Achievement in Preschool Children. *Early Education and*

Development, 25(1), 56-70. doi: 10.1080/10409289.2013.780505

Bennie, J. A., Teychenne, M. J., De Cocker, K., & Biddle, S. J. H. (2019). Associations between aerobic and muscle-strengthening exercise with depressive symptom severity among 17,839 U.S. adults. *Preventive Medicine, 121*, 121-127. doi: https://doi.org/10.1016/j.ypmed.2019.02.022

Berkel, C., Mauricio, A. M., Schoenfelder, E., & Sandler, I. N. (2011). Putting the Pieces Together: An Integrated Model of Program Implementation. *Prevention Science, 12*(1), 23-33. doi: 10.1007/s11121-010-0186-1

Biddle, S. J. H., Ciaccioni, S., Thomas, G., & Vergeer, I. (2019). Physical activity and mental health in children and adolescents: An updated review of reviews and an analysis of causality. *Psychology of Sport and Exercise, 42*, 146-155. doi: https://doi.org/10.1016/j.psychsport.2018.08.011

Biederman, J., Milberger, S., Faraone, S. V., Kiely, K., Guite, J., Mick, E., . . . Reed, E. (1995). Family-Environment Risk Factors for Attention-Deficit Hyperactivity Disorder: A Test of Rutter's Indicators of Adversity. *Archives of General Psychiatry, 52*(6), 464-470. doi: 10.1001/archpsyc.1995.03950180050007

Blair, S. N., Cheng, Y., & Holder, J. S. (2001). Is physical activity or physical fitness more important in defining health benefits? *Medicine & Science in Sports & Exercise, 33*(6, Suppl.), 379-399. doi: 10.1097/00005768-200106001-00007

Blashill, A. J., & Wilhelm, S. (2014). Body image distortions, weight, and depression in adolescent boys: Longitudinal trajectories into adulthood. *Psychology of men & masculinity, 15*(4), 445-451. doi: 10.1037/a0034618

Bonell, C., Fletcher, A., Morton, M., Lorenc, T., & Moore, L. (2012). Realist randomised controlled trials: A new approach to evaluating complex public health interventions. *Social Science & Medicine, 75*(12), 2299-2306. doi: 10.1016/j.socscimed.2012.08.032

Bonhauser, M., Fernandez, G., Püschel, K., Yañez, F., Montero, J., Thompson, B., & Coronado, G. (2005). Improving physical fitness and emotional well-being in adolescents of low socioeconomic status in Chile: results of a school-based

controlled trial. *Health Promotion International, 20*(2), 113-122. doi: 10.1093/heapro/dah603

Borgen, J. S., Gjølme, E. G., & Hallås, B. O. (2017). Kroppsøving blir redusert til "fysisk aktivitet" - debatten uteblir. [Physical education is reduced to «physical activity» - the debate is absent]. *Bedre Skole, 4*. Retrieved from https://episerver.utdanningsnytt.no/bedre-skole/debatt/2017/kroppsoving-er-mer-enn-fysisk-aktivitet/. [Norwegian]

Borgen, J. S., Moen, K. M., Gjølme, E. G., Løndal, K., & Hallås, B. O. (2018, March 22.). Mer fysisk aktivitet i skolen gir mindre læring. [More physical activity in school reduces learning]. *Bergens Tidende*. Retrieved from https://www.bt.no/btmeninger/debatt/i/WLewxK/mer-fysisk-aktivitet-i-skolen-gir-mindre-laering. [Norwegian]

Brattenborg, S., & Engebretsen, B. (2013). *Innføring i kroppsøvingsdidaktikk* (Vol. 3) [An introduction to physical education didactics]. Oslo: Cappelen Damm. [Norwegian]

Breistøl, S., Clench-Aas, J., Van Roy, B., & Raanaas, R. K. (2017). Association Between Participating in Noncompetitive or Competitive Sports and Mental Health among Adolescents – a Norwegian Population- based Cross- sectional Study. *Scandinavian Journal of Child and Adolescent Psychiatry and Psychology, 5*(1), 28-38. doi: 10.21307/sjcapp-2017-003

Bryan, H., & Burstow, B. (2018). Understanding Ethics in School-Based Research. *Professional Development in Education, 44*(1), 107-119. doi: 10.1080/19415257.2017.1361464

Button, T. M. M., Corley, R. P., Rhee, S. H., Hewitt, J. K., Young, S. E., & Stallings, M. C. (2007). Delinquent peer affiliation and conduct problems: A twin study. *Journal of Abnormal Psychology, 116*(3), 554-564. doi: 10.1037/0021-843X.116.3.554

Byrne, D. (2013). Evaluating complex social interventions in a complex world. *Evaluation, 19*(3), 217-228. doi: 10.1177/1356389013495617

Bøe, T., Hysing, M., Skogen, J., & Breivik, K. (2016). The Strengths and Difficulties Questionnaire (SDQ): Factor Structure and

Gender Equivalence in Norwegian Adolescents. *PLoS ONE, 11*(5), e0152202. doi: 10.1371/journal.pone.0152202

Bøe, T., Øverland, S., Lundervold, A. J., & Hysing, M. (2012). Socioeconomic status and children's mental health: results from the Bergen Child Study. *Social Psychiatry and Psychiatric Epidemiology, 47*(10), 1557-1566. doi: 10.1007/s00127-011-0462-9

Cairney, J., Kwan, M. Y. W., Velduizen, S., Hay, J., Bray, S. R., & Faught, B. E. (2012). Gender, perceived competence and the enjoyment of physical education in children: a longitudinal examination. *International Journal of Behavioral Nutrition and Physical Activity, 9*(1), 26. doi: 10.1186/1479-5868-9-26

Cardon, G., Philippaerts, R., Lefevre, J., Matton, L., Wijndaele, K., Balduck, A.-L., & Bourdeaudhuij, I. (2005). Physical Activity Levels In 10 To 11 Year-olds: Clustering Of Psycho-social Correlates. *Public Health Nutrition, 8*, 896-903. doi: 10.1079/PHN2005738

Casey, M. M., Harvey, J. T., Telford, A., Eime, R. M., Mooney, A., & Payne, W. R. (2014). Effectiveness of a school- community linked program on physical activity levels and health- related quality of life for adolescent girls. *BMC Public Health, 14*, 649-649. doi: 10.1186/1471-2458-14-649

Caspersen, C. J., Powell, K. E., & Christenson, G. M. (1985). Physical activity, exercise and physical fitness: Definitions and distinctions for health-related research. *Public Health Reports, 100*(2), 126-130. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1424733/pdf/pubhealthrep00100-0016.pdf

Cerin, E. (2010). Ways of unraveling how and why physical activity influences mental health through statistical mediation analyses. *Mental Health and Physical Activity, 3*(2), 51-60. doi: https://doi.org/10.1016/j.mhpa.2010.06.002

Christiansen, L. B., Lund-Cramer, P., Brondeel, R., Smedegaard, S., Holt, A.-D., & Skovgaard, T. (2018). Improving children's physical self-perception through a school-based physical activity intervention: The Move for Well-being in School study. *Mental Health and Physical Activity, 14*, 31-38. doi: 10.1016/j.mhpa.2017.12.005

Christie, D., & Viner, R. (2005). Adolescent development. *BMJ, 330*(7486), 301-304. doi: 10.1136/bmj.330.7486.301

Clarke, A., M. (2010). Context matters in programme implementation. *Health Education, 110*(4), 273-293. doi: 10.1108/09654281011052637

Collins, H., Booth, J., Duncan, A., Fawkner, S., & Niven, A. (2019). The Effect of Resistance Training Interventions on 'The Self' in Youth: a Systematic Review and Meta-analysis. *Sports Medicine - Open, 5*. doi: 10.1186/s40798-019-0205-0

Costigan, S. A., Lubans, D. R., Lonsdale, C., Sanders, T., & del Pozo Cruz, B. (2019). Associations between physical activity intensity and well-being in adolescents. *Preventive Medicine*. doi: https://doi.org/10.1016/j.ypmed.2019.05.009

Council of Europe (1988). *Handbook of the Eurofit tests of Physical Fitness*. Rome: Italian National Olympic Committee

Craft, A. W., & The Editors. (1998). The first randomised controlled trial. *Archives of Disease in Childhood 79*(5), 410-410. doi: 10.1136/adc.79.5.410

Craig, P., Dieppe, P., Macintyre, S., Michie, S., Nazareth, I., & Petticrew, M. (2008). Developing and evaluating complex interventions: the new Medical Research Council guidance. *BMJ, 337*, a1655. doi: 10.1136/bmj.a1655

Curtis, A. C. (2015). Defining adolescence. *Journal of Adolescent and Family Health, 7*(2), 2. Retrieved from https://scholar.utc.edu/cgi/viewcontent.cgi?article=1035&context=jafh

Dale, L. P., Vanderloo, L., Moore, S., & Faulkner, G. (2019). Physical activity and depression, anxiety, and self-esteem in children and youth: An umbrella systematic review. *Mental Health and Physical Activity, 16*, 66-79. doi: https://doi.org/10.1016/j.mhpa.2018.12.001

Dalene, K. E., Anderssen, S. A., Andersen, L. B., Steene-Johannessen, J., Ekelund, U., Hansen, B. H., & Kolle, E. (2018). Secular and longitudinal physical activity changes in population-based samples of children and adolescents. *Scandinavian journal of medicine & science in sports, 28*(1), 161-171. doi: 10.1111/sms.12876

Daly-Smith, A. J., Zwolinsky, S., McKenna, J., Tomporowski, P. D., Defeyter, M. A., & Manley, A. (2018). Systematic review of acute physically active learning and classroom movement breaks on children's physical activity, cognition, academic performance and classroom behaviour: understanding critical design features. *BMJ Open Sport & Exercise Medicine, 4*(1), e000341. doi: 10.1136/bmjsem-2018-000341

Deaton, A., & Cartwright, N. (2018). Understanding and misunderstanding randomized controlled trials. *Social Science & Medicine, 210*, 2-21. doi: https://doi.org/10.1016/j.socscimed.2017.12.005

DeSocio, J., & Hootman, J. (2004). Children's Mental Health and School Success. *Journal of School Nursing, 20*(4), 189-196. doi: 10.1177/10598405040200040201

Dobson, D., & Cook, T. J. (1980). Avoiding type III error in program evaluation: Results from a field experiment. *Evaluation and Program Planning, 3*(4), 269-276. doi: https://doi.org/10.1016/0149-7189(80)90042-7

Donenberg, G. R., & Pao, M. (2005). Youths and HIV/AIDS: Psychiatry's Role in a Changing Epidemic. *Journal of the American Academy of Child & Adolescent Psychiatry, 44*(8), 728-747. doi: https://doi.org/10.1097/01.chi.0000166381.68392.02

Dornbusch, S. M., Petersen, A. C., & Hetherington, E. M. (1991). Projecting the future of research on adolescence. *Journal of Research on Adolescence, 1*(1), 7-17.

Durlak, J. A. (2016). Programme implementation in social and emotional learning: basic issues and research findings. *Cambridge Journal of Education, 46*(3), 333-345. doi: 10.1080/0305764X.2016.1142504

Durlak, J. A., & Dupre, E. P. (2008). Implementation Matters: A Review of Research on the Influence of Implementation on Program Outcomes and the Factors Affecting Implementation. *American Journal of Community Psychology, 41*(3-4), 327-350. doi: 10.1007/s10464-008-9165-0

Dyrstad, S., Berg, T., & Tjelta, L. I. (2012). Secular trends in aerobic fitness performance in a cohort of Norwegian adolescents.

Scandinavian journal of medicine & science in sports, 22(6), 822-827. doi: https://doi.org/10.1111/j.1600-0838.2011.01315.x

Easterlin, M. C., Chung, P. J., Leng, M., & Dudovitz, R. (2019). Association of Team Sports Participation With Long-term Mental Health Outcomes Among Individuals Exposed to Adverse Childhood ExperiencesTeam Sports Participation and Long-term Mental Health in Individuals With Adverse Childhood ExperiencesTeam Sports Participation and Long-term Mental Health in Individuals With Adverse Childhood Experiences. *JAMA Pediatrics, 173*(7), 681-688. doi: 10.1001/jamapediatrics.2019.1212

Eather, N., Morgan, P. J., & Lubans, D. R. (2016). Effects of exercise on mental health outcomes in adolescents: Findings from the CrossFit™ teens randomized controlled trial. *Psychology of Sport and Exercise, 26,* 14-23. doi: https://doi.org/10.1016/j.psychsport.2016.05.008

Ebata, A. T., & Moos, R. H. (1991). Coping and adjustment in distressed and healthy adolescents. *Journal of Applied Developmental Psychology, 12*(1), 33-54. doi: https://doi.org/10.1016/0193-3973(91)90029-4

Edbom, T., Lichtenstein, P., Granlund, M., & Larsson, J.-O. (2006). Long-term relationships between symptoms of Attention Deficit Hyperactivity Disorder and self-esteem in a prospective longitudinal study of twins. *Acta Paediatrica, 95*(6), 650-657. doi: 10.1111/j.1651-2227.2006.tb02311.x

Eddolls, W. T. B., McNarry, M. A., Lester, L., Winn, C. O. N., Stratton, G., & Mackintosh, K. A. (2018). The association between physical activity, fitness and body mass index on mental well-being and quality of life in adolescents. *Quality of Life Research, 27*(9), 2313-2320. doi: 10.1007/s11136-018-1915-3

The Education Act (1998a) Act relating to Primary and Secondary Education and Training. (LOV-1998-07-17-61). Retrieved from https://lovdata.no/dokument/NLE/lov/1998-07-17-61#KAPITTEL_1

The Education Act (1998b) Act relating to Primary and Secondary Education and Training. (LOV-1998-07-17-61). Retrieved from

https://lovdata.no/dokument/NLE/lov/1998-07-17-61#KAPITTEL_1

Edwardson, C. L., Gorely, T., Pearson, N., & Atkin, A. (2013). Sources of activity-related social support and adolescents' objectively measured after-school and weekend physical activity: gender and age differences. *Journal of Physical Activity & Health, 10*(8), 1153-1158. doi: 10.1123/jpah.10.8.1153

Eime, R. M., Young, J. A., Harvey, J. T., Charity, M. J., & Payne, W. R. (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), 98. doi: 10.1186/1479-5868-10-98

Erola, J., Jalonen, S., & Lehti, H. (2016). Parental education, class and income over early life course and children's achievement. *Research in Social Stratification and Mobility, 44*, 33-43. doi: https://doi.org/10.1016/j.rssm.2016.01.003

Finlay, L. (2002). "Outing" the Researcher: The Provenance, Process, and Practice of Reflexivity. *Qualitative Health Research, 12*(4), 531-545. doi: 10.1177/104973202129120052

Fixsen, D. L., Naoom, S. F., Blase, K. A., Friedman, R. M., & Wallace, F. (2005). *Implementation research : a synthesis of the literature.* Tampa, Florida: University of South Florida. FMHI Publication 231. Retrieved from: https://nirn.fpg.unc.edu/sites/nirn.fpg.unc.edu/files/resources/NIRN-MonographFull-01-2005.pdf

Foster, C. (1995). Why do research ethics committees disagree with each other? *Journal of the Royal College of Physicians of London, 29*(4), 315-318

Frey, J. H., & Fontana, A. (1991). The group interview in social research. *The Social Science Journal, 28*(2), 175-187. doi: https://doi.org/10.1016/0362-3319(91)90003-M

Fullan, M. (2016). *The New meaning of educational change* (Fifth edition. ed.). London: Routledge.

Gapin, J. I., Labban, J. D., & Etnier, J. L. (2011). The effects of physical activity on attention deficit hyperactivity disorder symptoms: The evidence. *Preventive Medicine, 52*, S70-S74. doi: https://doi.org/10.1016/j.ypmed.2011.01.022

Garnefski, N., & Diekstra, R. F. W. (1996). Perceived Social Support from Family, School, and Peers: Relationship with Emotional and Behavioral Problems among Adolescents. *Journal of the American Academy of Child & Adolescent Psychiatry*, *35*(12), 1657-1664. doi: https://doi.org/10.1097/00004583-199612000-00018

Gerber, M., Kalak, N., Lemola, S., Clough, P. J., Pühse, U., Elliot, C., . . . Brand, S. (2012). Adolescents' exercise and physical activity are associated with mental toughness. *Mental Health and Physical Activity*, *5*(1), 35-42. doi: https://doi.org/10.1016/j.mhpa.2012.02.004

Gerber, P. (2006). What can we learn from the Hwang and Sudbø affairs? *Medical Journal of Australia*, *184*(12), 632-635. doi: 10.5694/j.1326-5377.2006.tb00420.x

Goksøyr, M. (2007). «Vi skal ikke være rævasittere i et departement». Rolf Hofmos visjoner og virke for folkeidretten. ["We are not going to be sitting on our asses in a ministry". Rolf Hofmo's visions and work for people's sports] In *Arbeiderhistorie 2007* (Vol. 1, pp. 33-61). Oslo: Forlaget Aktuell AS. [Norwegian]

Goodman, A., & Goodman, R. (2009). Strengths and Difficulties Questionnaire as a Dimensional Measure of Child Mental Health. *Journal of the American Academy of Child & Adolescent Psychiatry*, *48*(4), 400-403. doi: https://doi.org/10.1097/CHI.0b013e3181985068

Goodman, R. (1997). The Strengths and Difficulties Questionnaire: A Research Note. *Journal of Child Psychology and Psychiatry*, *38*(5), 581-586. doi: 10.1111/j.1469-7610.1997.tb01545.x

Goodman, R. (2001). Psychometric Properties of the Strengths and Difficulties Questionnaire. *Journal of the American Academy of Child & Adolescent Psychiatry*, *40*(11), 1337-1345. doi: https://doi.org/10.1097/00004583-200111000-00015

Goodman, R., Ford, T., Simmons, H., Gatward, R., & Meltzer, H. (2000). Using the Strengths and Difficulties Questionnaire (SDQ) to screen for child psychiatric disorders in a community sample. *The British journal of psychiatry : the journal of mental science*, *177*, 534. doi: https://doi.org/10.1192/bjp.177.6.534

Goodman, R., Meltzer, H., & Bailey, V. (2003). The Strengths and Difficulties Questionnaire: a pilot study on the validity of the

self-report version. *International Review of Psychiatry, 15*(1-2), 173-177. doi: 10.1080/0954026021000046137
Grabe, S., Ward, L. M., & Hyde, J. S. (2008). The role of the media in body image concerns among women: A meta-analysis of experimental and correlational studies. *Psychological Bulletin, 134*(3), 460-476. doi: 10.1037/0033-2909.134.3.460
Greenland, S., Senn, S. J., Rothman, K. J., Carlin, J. B., Poole, C., Goodman, S. N., & Altman, D. G. (2016). Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations. *European Journal of Epidemiology, 31*(4), 337-350. doi: 10.1007/s10654-016-0149-3
Greenleaf, C., Petrie, T., & Martin, S. B. (2010). Psychosocial Variables Associated With Body Composition and Cardiorespiratory Fitness in Middle School Students. *Research Quarterly for Exercise and Sport, 81,* 65-74. doi: 10.1080/02701367.2010.10599695
Ha, C., Petersen, N., & Sharp, C. (2008). Narcissism, self-esteem, and conduct problems. *European child & adolescent psychiatry, 17*(7), 406-413. doi: 10.1007/s00787-008-0682-z
Hamilton, D. (1990). Publishing by – and for? – the Numbers. *Science, 250*(4986), 1331-1332. doi: 10.1126/science.2255902
Hammersley, M. (1987). Some Notes on the Terms 'Validity' and 'Reliability'. *British Educational Research Journal, 13*(1), 73-82. doi: 10.1080/0141192870130107
Hammersley, M. (1992). *What's wrong with ethnography?* London and New York: Routledge.
Harter, S. (2003). The development of self-representations during childhood and adolescence. In M. R. Leary & J. P. Tangney (eds.) *Handbook of self and identity* (pp. 610-642). New York: The Guilford Press
Harter, S., & Whitesell, N. R. (2001). On the importance of importance ratings in understanding adolescents' self-esteem: Beyond statistical parsimony. In R. J. Riding & S. G. Rayner (eds.) *International perspectives on individual differnces, Vol. 2. Self perception* (pp. 3-23). London: Ablex Publishing
Harvey, S. B., Øverland, S., Hatch, S. L., Wessely, S., Mykletun, A., & Hotopf, M. (2017). Exercise and the Prevention of Depression:

Results of the HUNT Cohort Study. *The American journal of psychiatry*, *175*(1) doi: 10.1176/appi.ajp.2017.16111223
Haugen, T., Høigaard, R., & Seiler, S. (2014). Normative data of BMI and physical fitness in a Norwegian sample of early adolescents. *Scandinavian Journal of Public Health*, *42*(1), 67-73. doi: 10.1177/1403494813504502
Haugen, T., Säfvenbom, R., & Ommundsen, Y. (2011). Physical activity and global self-worth: The role of physical self-esteem indices and gender. *Mental Health and Physical Activity*, *4*(2), 49-56. doi: https://doi.org/10.1016/j.mhpa.2011.07.001
Hawe, P. (2015). Lessons from Complex Interventions to Improve Health. *Annual Review of Public Health*, *36*(1), 307-323. doi: 10.1146/annurev-publhealth-031912-114421
Hawe, P., Shiell, A., & Riley, T. (2004). Complex interventions: how "out of control" can a randomised controlled trial be? *BMJ*, *328*(7455), 1561-1563. doi: 10.1136/bmj.328.7455.1561
Hawe, P., Shiell, A., & Riley, T. (2009). Theorising Interventions as Events in Systems. *American Journal of Community Psychology*, *43*(3-4), 267-276. doi: 10.1007/s10464-009-9229-9
Heelan, K., Welk, G., Pfeiffer Karin, A., Eisenmann Joey, C., Drenowatz, C., Gentile, D., & Walsh, D. (2010). Influence of socio-economic status on habitual physical activity and sedentary behavior in 8- to 11-year old children. *BMC Public Health*, *10*(1), 214. doi: 10.1186/1471-2458-10-214
Hofmann, B., Myhr, A. I., & Holm, S. (2013). Scientific dishonesty—a nationwide survey of doctoral students in Norway. *BMC Medical Ethics*, *14*(1), 3. doi: 10.1186/1472-6939-14-3
Hughes, R. A., Heron, J., Sterne, J. A. C., & Tilling, K. (2019). Accounting for missing data in statistical analyses: multiple imputation is not always the answer. *International Journal of Epidemiology*, *48*(4), 1294-1304. doi: 10.1093/ije/dyz032
Janssen, A., Leahy, A. A., Diallo, T. M. O., Smith, J. J., Kennedy, S. G., Eather, N., . . . Lubans, D. R. (2020). Cardiorespiratory fitness, muscular fitness and mental health in older adolescents: A multi-level cross-sectional analysis. *Preventive Medicine*, *132*, 105985. doi: https://doi.org/10.1016/j.ypmed.2020.105985
Jarvis, P. (2006). *Towards a Comprehensive Theory of Human Learning*. New York, USA: Routledge.

Jetté, M., Sidney, K., & Blümchen, G. (1990). Metabolic equivalents (METS) in exercise testing, exercise prescription, and evaluation of functional capacity. *Clinical Cardiology, 13*(8), 555-565. doi: 10.1002/clc.4960130809

Kamp, C. F., Sperlich, B., & Holmberg, H.-C. (2014). Exercise reduces the symptoms of attention-deficit/hyperactivity disorder and improves social behaviour, motor skills, strength and neuropsychological parameters. *Acta Paediatrica, 103*(7), 709-714. doi: 10.1111/apa.12628

Kandola, A., Lewis, G., Osborn, D. P. J., Stubbs, B., & Hayes, J. F. (2020). Depressive symptoms and objectively measured physical activity and sedentary behaviour throughout adolescence: a prospective cohort study. *The Lancet Psychiatry, 7*(3), 262-271. doi: 10.1016/S2215-0366(20)30034-1

Keenan-Miller, D., Hammen, C. L., & Brennan, P. A. (2007). Health Outcomes Related to Early Adolescent Depression. *Journal of Adolescent Health, 41*(3), 256-262. doi: https://doi.org/10.1016/j.jadohealth.2007.03.015

Kelly, N. R., Mazzeo, S. E., Evans, R. K., Stern, M., Thacker, L. F., Thornton, L. M., & Laver, J. H. (2010). Physical activity, fitness and psychosocial functioning of obese adolescents. *Mental Health and Physical Activity, 4*(1), 31-37. doi: https://doi.org/10.1016/j.mhpa.2010.11.001

Keshavarz, N., Nutbeam, D., Rowling, L., & Khavarpour, F. (2010). Schools as social complex adaptive systems: A new way to understand the challenges of introducing the health promoting schools concept. *Social Science & Medicine, 70*(10), 1467-1474. doi: https://doi.org/10.1016/j.socscimed.2010.01.034

Kieling, C., Baker-Henningham, H., Belfer, M., Conti, G., Ertem, I., Omigdobun, O., . . . Rahman, A. (2011). Child and adolescent mental health worldwide: evidence for action. *Lancet, 378*, 1515-1525. doi: 10.1016/S01406736(11)60827-1

Kinge, J. M., Sælensminde, K., Dieleman, J., Vollset, S. E., & Norheim, O. F. (2017). Economic losses and burden of disease by medical conditions in Norway. *Health Policy, 121*(6), 691-698. doi: https://doi.org/10.1016/j.healthpol.2017.03.020

Kirke og undervisningsdepartementet. [Ministry of Church Affairs and Education] (1974). *Mønsterplan for grunnskolen: bokmål*.

[Curriculum plan for primary and lower secondary school] Oslo: Aschehoug. [Norwegian]

Kirke og undervisningsdepartementet. [Ministry of Church Affairs and Education] (1987). *Mønsterplan for grunnskolen: M87*. [Curriculum plan for primary and lower secondary school] Oslo: Aschehoug. [Norwegian]

Kirke, undervisnings og forskningsdepartementet [Ministry of Church Affairs, Education and Research] (1996). *Læreplanverket for den 10-årige grunnskolen*. Oslo: Nasjonalt læremiddelsenter. [Norwegian]

Kolle, E., Støren Stokke, J., H Hansen, B., & Anderssen, S. (2012). *Fysisk aktivitet blant 6-, 9- og 15-åringer i Norge - Resultater fra en kartlegging i 2011*. [Physical activity among 6-, 9-, and 15-year-olds in Norway – Results from an investigation in 2011] Retrieved from https://helsedirektoratet.no/Lists/Publikasjoner/Attachments/710/Fysisk-aktivitet-blant-%206-9-og-15-aringer-i-norge-resultater-fra-en-kartlegging-i-2011-IS-2002.pdf

Kolle, E., Solberg, R. B., Säfvenbom, R., Dyrstad, S. M., Berntsen, S., Ekelund, U., . . . Grydeland, M. (2020). Change in physical activity and aerobic fitness after a school-based intervention among adolescents: The school in motion cluster randomized trial *Manuscript submitted for publication*.

Koskelainen, M., Sourander, A., & Vauras, M. (2001). Self-reported strengths and difficulties in a community sample of Finnish adolescents. *Journal of European Child & Adolescent Psychiatry, 10*(3), 180-185. doi:10.1007/s007870170024

La Greca, A. M., & Lopez, N. (1998). Social Anxiety Among Adolescents: Linkages with Peer Relations and Friendships. *Journal of Abnormal Child Psychology, 26*(2), 83-94. doi: 10.1023/A:1022684520514

Laberge, S., Bush, P. L., & Chagnon, M. (2012). Effects of a Culturally Tailored Physical Activity Promotion Program on Selected Self-Regulation Skills and Attitudes in Adolescents of an Underserved, Multiethnic Milieu. *American Journal of Health Promotion, 26*(4), e105-e115. doi: 10.4278/ajhp.090625-QUAN-202

Lakes, K. D., & Hoyt, W. T. (2004). Promoting self-regulation through school-based martial arts training. *Journal of Applied Developmental Psychology, 25*(3), 283-302. doi: https://doi.org/10.1016/j.appdev.2004.04.002

Lamb, D., & Gulliford, A. (2011). Physical exercise and children's self-concept of emotional and behavioural well-being: A randomised controlled trial. *Educational & Child Psychology, 28*(4), 66-74.

Lang, C., Brand, S., Feldmeth, A. K., Holsboer-Trachsler, E., Pühse, U., & Gerber, M. (2013). Increased self-reported and objectively assessed physical activity predict sleep quality among adolescents. *Physiology & Behavior, 120*, 46-53. doi: https://doi.org/10.1016/j.physbeh.2013.07.001

Lawler, M., & Nixon, E. (2011). Body Dissatisfaction Among Adolescent Boys and Girls: The Effects of Body Mass, Peer Appearance Culture and Internalization of Appearance Ideals. *Journal of Youth and Adolescence, 40*(1), 59-71. doi: 10.1007/s10964-009-9500-2

Laxdal, A., & Giske, R. (2019). Gender and the perceived learning environment in upper secondary school physical education. *Sport, Education and Society*, 1-9. doi: 10.1080/13573322.2019.1666360

Lee, K. J., & Carlin, J. B. (2012). Recovery of information from multiple imputation: a simulation study. *Emerging Themes in Epidemiology, 9*(1), 3. doi: 10.1186/1742-7622-9-3

Leising, D., Locke, K. D., Kurzius, E., & Zimmermann, J. (2016). Quantifying the Association of Self-Enhancement Bias With Self-Ratings of Personality and Life Satisfaction. *Assessment, 23*(5), 588-602. doi: 10.1177/1073191115590852

Lerner, R. M., Hershberg, R., Hilliard, L., & Johnson, S. (2015). Concepts and Theories in Human Development. In M. H. Bornstein & M. E. Lamb (Eds.), *Developmental science: An advanced textbook* (pp. 3-42). New York: Psychology Press.

Lerner, R. M., & Lerner, J. V. (2013). *The Positive Development of Youth: Comprehensive findings from the 4-H study of positive youth development*. Washington: Tufts University. Retrieved from: https://4-h.org/wp-content/uploads/2016/02/4-H-Study-of-Positive-Youth-Development-Full-Report.pdf

Lin, K., Stubbs, B., Zou, W., Zheng, W., Lu, W., Gao, Y., . . . Xu, G. (2020). Aerobic exercise impacts the anterior cingulate cortex in adolescents with subthreshold mood syndromes: a randomized controlled trial study. *Translational Psychiatry, 10*(1), 155. doi: 10.1038/s41398-020-0840-8

Lott, M. A., & Jensen, C. D. (2016). Executive Control Mediates the Association Between Aerobic Fitness and Emotion Regulation in Preadolescent Children. *Journal of Pediatric Psychology, 42*(2), 162-173. doi: 10.1093/jpepsy/jsw052

Lubans, D. R., & Cliff, D. P. (2011). Muscular fitness, body composition and physical self-perception in adolescents. *Journal of Science and Medicine in Sport, 14*(3), 216-221. doi: https://doi.org/10.1016/j.jsams.2010.10.003

Lubans, D. R., Richards, J., Hillman, C., Faulkner, G., Beauchamp, M., Nilsson, M., . . . Biddle, S. (2016). Physical Activity for Cognitive and Mental Health in Youth: A Systematic Review of Mechanisms. *Pediatrics, 138*(3), e20161642. doi:10.1542/peds.2016-1642

Lubans, D. R., Smith, J. J., Morgan, P. J., Beauchamp, M. R., Miller, A., Lonsdale, C., . . . Dally, K. (2016). Mediators of Psychological Well-being in Adolescent Boys. *Journal of Adolescent Health, 58*(2), 230-236. doi: 10.1016/j.jadohealth.2015.10.010

Ludvigsen, S., Elverhøi, P., Ishaq, B., Rasmussen, J., Sundberg, D., Gundersen, E., . . . Rose, S. (2014). *Elevenes læring i fremtidens skole*. [Students' learning in future's school]. Retrieved from Oslo: https://www.regjeringen.no/contentassets/e22a715fa374474581a8c58288edc161/no/pdfs/nou201420140007000dddpdfs.pdf [Norwegian]

Luppino, F. S., de Wit, L. M., Bouvy, P. F., Stijnen, T., Cuijpers, P., Phennix, B. W. J. H., & Zitman, F. G. (2010). Overweight, obesity, and depression: A systematic review and meta-analysis of longitudinal studies. *Archives of General Psychiatry, 67*(3), 220-229. doi: 10.1001/archgenpsychiatry.2010.2

Malecki, C., & Demaray, M. (2002). Measuring perceived social support: Development of the child and adolescent social support

scale (CASSS). *Psychology in the Schools, 39*, 1-18. doi: 10.1002/pits.10004
Manly, C. A., & Wells, R. S. (2015). Reporting the Use of Multiple Imputation for Missing Data in Higher Education Research. *Research in Higher Education, 56*(4), 397-409. doi: 10.1007/s11162-014-9344-9
Manzano, A. (2016). The craft of interviewing in realist evaluation. *Evaluation, 22*(3), 342-360. doi: 10.1177/1356389016638615
Markey, C. N. (2010). Invited Commentary: Why Body Image is Important to Adolescent Development. *Journal of Youth and Adolescence, 39*(12), 1387-1391. doi:10.1007/s10964-010-9510-0
Marmorstein, N. R., & Iacono, W. G. (2004). Major depression and conduct disorder in youth: associations with parental psychopathology and parent–child conflict. *Journal of Child Psychology and Psychiatry, 45*(2), 377-386. doi: 10.1111/j.1469-7610.2004.00228.x
Mason, M. (2008). What Is Complexity Theory and What Are Its Implications for Educational Change? *Educational Philosophy and Theory, 40*(1), 35-49. doi: 10.1111/j.1469-5812.2007.00413.x
Mastoras, S., Saklofske, D., Schwean, V., & Climie, E. (2015). Social Support in Children With ADHD: An Exploration of Resilience. *Journal of attention disorders, 22*. doi: 10.1177/1087054715611491
Matta Mello Portugal, E., Cevada, T., Sobral Monteiro-Junior, R., Teixeira Guimarães, T., da Cruz Rubini, E., Lattari, E., . . . Camaz Deslandes, A. (2013). Neuroscience of Exercise: From Neurobiology Mechanisms to Mental Health. *Neuropsychobiology, 68*(1), 1-14. doi: 10.1159/000350946
Mays, N., & Pope, C. (1995). Qualitative Research: Observational methods in health care settings. *BMJ, 311*(6998), 182-184. doi: 10.1136/bmj.311.6998.182
McClelland, M. M., & Cameron, C. E. (2011). Self-regulation and academic achievement in elementary school children. *New Directions for Child and Adolescent Development, 2011*(133), 29-44. doi: 10.1002/cd.302

McShane, B. B., Gal, D., Gelman, A., Robert, C., & Tackett, J. L. (2019). Abandon Statistical Significance. *The American Statistician, 73*(sup1), 235-245. doi: 10.1080/00031305.2018.1527253

Meteyard, L., & Davies, R. A. I. (2020). Best practice guidance for linear mixed-effects models in psychological science. *Journal of Memory and Language, 112*, 104092. doi: https://doi.org/10.1016/j.jml.2020.104092

Ministry of Health and Care Services. (2003). *Report No. 16 to the Storting (2002-2003)*. Retrieved from https://www.regjeringen.no/en/dokumenter/report-no.-16-to-the-storting-2002-2003/id196640/

Ministry of Health and Care Services. (2004). *Action plan on physical activity 2005 - 2009*. Retrieved from https://www.regjeringen.no/en/dokumenter/action-plan-on-physical-activity-2005---/id102065/

Ministry of Health and Care Services. (2015). *Stortingsmelding 19: Folkehelsemeldingen - Mestring og muligheter*. [Report No. 19 to the Storting: The public health message – mastery and opportunity] Oslo: Regjeringen [Norwegian]

Moksnes, U. K., & Espnes, G. A. (2012). Self-esteem and emotional health in adolescents – gender and age as potential moderators. *Scandinavian Journal of Psychology, 53*(6), 483-489. doi: 10.1111/sjop.12021

Moksnes, U. K., Moljord, I. E. O., Espnes, G. A., & Byrne, D. G. (2010). The association between stress and emotional states in adolescents: The role of gender and self-esteem. *Personality and Individual Differences, 49*(5), 430-435. doi: https://doi.org/10.1016/j.paid.2010.04.012

Moliner-Urdiales, D., Ruiz, J. R., Vicente-Rodriguez, G., Ortega, F. B., Rey-Lopez, J. P., España-Romero, V., . . . Moreno, L. A. (2011). Associations of muscular and cardiorespiratory fitness with total and central body fat in adolescents: The HELENA Study. *British Journal of Sports Medicine 45*(2), 101-108. doi: 10.1136/bjsm.2009.062430

Moore, G., Raisanen, L., Moore, L., Din, N. U., & Murphy, S. (2013). Mixed-method process evaluation of the Welsh National

Exercise Referral Scheme. *Health Education, 113*. doi: 10.1108/HE-08-2012-0046

Moore, G. F., Audrey, S., Barker, M., Bond, L., Bonell, C., Hardeman, W., . . . Baird, J. (2015). Process evaluation of complex interventions: Medical Research Council guidance. *British Medical Journal, 350*. doi: 10.1136/bmj.h1258

Moore, G. F., Evans, R. E., Hawkins, J., Littlecott, H., Melendez-Torres, G. J., Bonell, C., & Murphy, S. (2019). From complex social interventions to interventions in complex social systems: Future directions and unresolved questions for intervention development and evaluation. *Evaluation, 25*(1), 23-45. doi: 10.1177/1356389018803219

Moore, J., Bumbarger, B., & Cooper, B. (2013). Examining Adaptations of Evidence-Based Programs in Natural Contexts. *The journal of primary prevention, 34*(3). doi: 10.1007/s10935-013-0303-6

Moore, R. D., Drollette, E. S., Scudder, M. R., Bharij, A., & Hillman, C. H. (2014). The influence of cardiorespiratory fitness on strategic, behavioral, and electrophysiological indices of arithmetic cognition in preadolescent children. *Frontiers in Human Neuroscience, 8*(258). doi: 10.3389/fnhum.2014.00258

Murberg, T. A., & Bru, E. (2004). School-Related Stress and Psychosomatic Symptoms among Norwegian Adolescents. *25*(3), 317-332. doi: 10.1177/0143034304046904

Muris, P. (2002). Relationships between self-efficacy and symptoms of anxiety disorders and depression in a normal adolescent sample. *Personality and Individual Differences, 32*(2), 337-348. doi: https://doi.org/10.1016/S0191-8869(01)00027-7

Muris, P., Meesters, C., Eijkelenboom, A., & Vincken, M. (2004). The self-report version of the Strengths and Difficulties Questionnaire: Its psychometric properties in 8- to 13-year-old non-clinical children. *British. Journal of Clinical Psychology, 43*, 437-448. doi: 10.1348/0144665042388982

Mykletun, A., Lydersen, S., Bjørnelv, S., & Holmen Turid, L. (2007). Changes in BMI-distribution from 1966–69 to 1995–97 in adolescents. The Young-HUNT study, Norway. *BMC Public Health, 7*(1), 279. doi: 10.1186/1471-2458-7-279

Nardini, C. (2014). The ethics of clinical trials. *Ecancermedicalscience, 8*, 387-387. doi: 10.3332/ecancer.2014.387

Naylor, P.-J., Nettlefold, L., Race, D., Hoy, C., Ashe, M. C., Wharf Higgins, J., & McKay, H. A. (2015). Implementation of school based physical activity interventions: A systematic review. *Preventive Medicine, 72*, 95-115. doi: 10.1016/j.ypmed.2014.12.034

Nestler, E. J., Barrot, M., DiLeone, R. J., Eisch, A. J., Gold, S. J., & Monteggia, L. M. (2002). Neurobiology of Depression. *Neuron, 34*(1), 13-25. doi: https://doi.org/10.1016/S0896-6273(02)00653-0

Ng, Q. X., Ho, C. Y. X., Chan, H. W., Yong, B. Z. J., & Yeo, W.-S. (2017). Managing childhood and adolescent attention-deficit/hyperactivity disorder (ADHD) with exercise: A systematic review. *Complementary Therapies in Medicine, 34*, 123-128. doi: https://doi.org/10.1016/j.ctim.2017.08.018

Norris, R., Carroll, D., & Cochrane, R. (1992). The effects of physical activity and exercise training on psychological stress and well-being in an adolescent population. *Journal of Psychosomatic Research, 36*(1), 55-65. doi: 10.1016/0022-3999(92)90114-H

The Norwegian Directorate for Education and Training. (n. d.) *Core curriculum: Health and life skills.* Retrieved from https://www.udir.no/lk20/overordnet-del/prinsipper-for-laring-utvikling-og-danning/tverrfaglige-temaer/folkehelse-og-livsmestring/?lang=eng

The Norwegian Directorate for Education and Training. (2015). *Curriculum for physical education (KRO1-04)*. Oslo: The Norwegian Directorate for Education and Training. Retrieved from https://www.udir.no/kl06/KRO1-04/Hele/Formaal?lplang=http://data.udir.no/kl06/eng

The Norwegian Directorate for Education and Training. (2012). *Curriculum for the elective subject of Physical Activity and Health (FAH1-01)*. Oslo Retrieved from https://www.udir.no/kl06/FAH1-01?lplang=http://data.udir.no/kl06/eng

The Norwegian National Research Ethics Committees (2016). *Guidelines for Research Ethics in the Social Sciences, Humanities, Law and Theology*. In The Norwegian National

Research Ethics Committees (Ed.), (pp. 44). Retrieved from https://www.etikkom.no/globalassets/documents/english-publications/60127_fek_guidelines_nesh_digital_corr.pdf

Oakley, A., Strange, V., Bonell, C., Allen, E., & Stephenson, J. (2006). Process evaluation in randomised controlled trials of complex interventions. *BMJ, 332*(7538), 413-416. doi: 10.1136/bmj.332.7538.413

Okado, Y., & Bierman, K. L. (2015). Differential Risk for Late Adolescent Conduct Problems and Mood Dysregulation Among Children with Early Externalizing Behavior Problems. *Journal of Abnormal Child Psychology, 43*(4), 735-747. doi: 10.1007/s10802-014-9931-4

Olive, L. S., Byrne, D. G., Cunningham, R. B., & Telford, R. D. (2012). Effects of physical activity, fitness and fatness on children's body image: The Australian LOOK longitudinal study. *Mental Health and Physical Activity, 5*(2), 116-124. doi: https://doi.org/10.1016/j.mhpa.2012.08.004

Oppedal, B., & Røysamb, E. (2004). Mental health, life stress and social support among young Norwegian adolescents with immigrant and host national background. *Scandinavian Journal of Psychology, 45*(2), 131-144. doi: 10.1111/j.1467-9450.2004.00388.x

Oppedal, B., Røysamb, E., & Heyerdahl, S. (2005). Ethnic group, acculturation, and psychiatric problems in young immigrants. *Journal of Child Psychology and Psychiatry, 46*(6), 646-660. doi: 10.1111/j.1469-7610.2004.00381.x

Padilla-Moledo, C., Ruiz, J. R., Ortega, F. B., Mora, J., & Castro-Pinero, J. (2012). Associations of muscular fitness with psychological positive health, health complaints, and health risk behaviors in Spanish children and adolescents. *Journal of Strength and Conditioning Research, 26*(1), 167-173. doi: 10.1519/JSC.0b013e31821c2433

Pate, R. R., Pratt, M., Blair, S. N., Haskell, W. L., Macera, C. A., Bouchard, C., . . . Wilmore, J. H. (1995). Physical Activity and Public Health: A recommendation from the Centers for Disease Control and Prevention and the American College of Sports Medicine. *Journal of American Medical Association, 273*, 402-407. doi: 10.1097/00043764-199606000-00005

Patel, V., Flisher, A. J., Hetrick, S., & McGorry, P. (2007). Mental health of young people: a global public- health challenge. *Lancet, 369*(9569), 1302. doi: 10.1016/S0140-6736(07)60368-7

Pedersen, A. B., Mikkelsen, E. M., Cronin-Fenton, D., Kristensen, N. R., Pham, T. M., Pedersen, L., & Petersen, I. (2017). Missing data and multiple imputation in clinical epidemiological research. *Clinical Epidemiology, 9,* 157-166. doi: 10.2147/CLEP.S129785

Pels, F., & Kleinert, J. (2016). Loneliness and physical activity: A systematic review. *International Review of Sport and Exercise Psychology, 9*, 1-30. doi: 10.1080/1750984X.2016.1177849

Pereira-Miranda, E., Costa, P. R., Queiroz, V. A., Pereira-Santos, M., & Santana, M. L. (2017). Overweight and obesity associated with higher depression prevalence in adults: a systematic review and meta-analysis. *Journal of the American College of Nutrition, 36*(3), 223-233.

Pinto Pereira, S. M., Geoffroy, M., & Power, C. (2014). Depressive symptoms and physical activity during 3 decades in adult life: Bidirectional associations in a prospective cohort study. *JAMA Psychiatry, 71*(12), 1373-1380. doi: 10.1001/jamapsychiatry.2014.1240

Prinstein, M. J., Borelli, J. L., Cheah, C. S. L., Simon, V. A., & Aikins, J. W. (2005). Adolescent Girls' Interpersonal Vulnerability to Depressive Symptoms: A Longitudinal Examination of Reassurance-Seeking and Peer Relationships. *Journal of Abnormal Psychology, 114*(4), 676-688. doi: 10.1037/0021-843X.114.4.676

Puskar, K., Sereika, S., & Tusaie-Mumford, K. (2003). Effect of the Teaching Kids to Cope (TKC©) Program on Outcomes of Depression and Coping Among Rural Adolescents. *Journal of Child and Adolescent Psychiatric Nursing, 16*(2), 71-80. doi: 10.1111/j.1744-6171.2003.tb00350.x

Raffaelli, M., Crockett, L. J., & Shen, Y.-L. (2005). Developmental Stability and Change in Self-Regulation From Childhood to Adolescence. *The Journal of Genetic Psychology, 166*(1), 54-76. doi: 10.3200/GNTP.166.1.54-76

Ragelienė, T. (2016). Links of Adolescents Identity Development and Relationship with Peers: A Systematic Literature Review.

Journal of the Canadian Academy of Child and Adolescent Psychiatry, *25*(2), 97-105.

Resaland, G. K. (2018, January 24.). Ein time dagleg fysisk aktivitet i skulen er mogleg å få til. [One hour daily physical activity in school is possible] *VG*. Retrieved from https://www.vg.no/nyheter/meninger/i/8wBoJE/ein-time-dagleg-fysisk-aktivitet-i-skulen-er-mogleg-aa-faa-til. [Norwegian]

Revicki, D. A., & Frank, L. (1999). Pharmacoeconomic Evaluation in the Real World. *PharmacoEconomics*, *15*(5), 423-434. doi: 10.2165/00019053-199915050-00001

Rieck, T., Jackson, A., Martin, S. B., Petrie, T., & Greenleaf, C. (2013). Health-related Fitness, Body Mass Index, and Risk of Depression among Adolescents. *Medicine & Science in Sports & Exercise*, *45*(6), 1083-1088. doi: 10.1249/MSS.0b013e3182831db1

Rimm-Kaufman, S. E., Fan, X., Chiu, Y.-J., & You, W. (2007). The contribution of the Responsive Classroom Approach on children's academic achievement: Results from a three year longitudinal study. *Journal of School Psychology*, *45*(4), 401-421. doi: https://doi.org/10.1016/j.jsp.2006.10.003

Riso, D. D., Salcuni, S., Chessa, D., Raudino, A., Lis, A., & Altoè, G. (2010). The Strengths and Difficulties Questionnaire (SDQ). Early evidence of its reliability and validity in a community sample of Italian children. *Personality and Individual Differences*, *49*(6), 570-575. doi: https://doi.org/10.1016/j.paid.2010.05.005

Robinson, O. C. (2014). Sampling in Interview-Based Qualitative Research: A Theoretical and Practical Guide. *Qualitative Research in Psychology*, *11*(1), 25-41. doi: 10.1080/14780887.2013.801543

Rogers, E. M. (1995). *Diffusion of innovations* (4th ed. ed.). New York: Free Press.

Rosenbaum, S., Tiedemann, A., Sherrington, C., Curtis, J., & Ward, P. (2014). Physical Activity Interventions for People With Mental Illness: A Systematic Review and Meta- Analysis. In *J. Clin. Psychiatry* (Vol. 75, pp. 964-U171).

Rothenberger, A., Becker, A., Erhart, M., Wille, N., & Ravens-Sieberer, U. (2008). Psychometric properties of the parent strengths and difficulties questionnaire in the general population of German children and adolescents: results of the BELLA study. *European child & adolescent psychiatry, 17*(1), 99. doi: 10.1007/s00787-008-1011-2

Ruchkin, V., Jones, S., Vermeiren, R., & Schwab-Stone, M. (2008). The strengths and difficulties questionnaire: The self-report version in American urban and suburban youth. *Psychological Assessment, 20*(2), 175-182. doi: 10.1037/1040-3590.20.2.175

Ruggero, C. J., Petrie, T., Sheinbein, S., Greenleaf, C., & Martin, S. (2015). Cardiorespiratory Fitness May Help in Protecting Against Depression Among Middle School Adolescents. *Journal of Adolescent Health, 57*(1), 60-65. doi: https://doi.org/10.1016/j.jadohealth.2015.03.016

Rutter, H., Savona, N., Glonti, K., Bibby, J., Cummins, S., Finegood, D. T., . . . White, M. (2017). The need for a complex systems model of evidence for public health. *The Lancet, 390*(10112), 2602-2604. doi: 10.1016/S0140-6736(17)31267-9

Ryan, R. M., & Deci, E. L. (2002). An overview of Self-determination Theory: An organismic-dialectical perspective. In E. L. Deci & R. M. Ryan (Eds.), *Handbook of Self-determination Research* (pp. 3-33). Rochester, New York: The University of Rochester Press.

Sagatun, A., Søgaard, A. J., Bjertness, E., Selmer, R., & Heyerdahl, S. (2007). The association between weekly hours of physical activity and mental health: A three-year follow-up study of 15–16-year-old students in the city of Oslo, Norway. *BMC Public Health, 7*(1). doi: 10.1186/1471-2458-7-155

Sagatun, Å., Kolle, E., Anderssen, S. A., Thoresen, M., & Søgaard, A. J. (2008). Three-year follow-up of physical activity in Norwegian youth from two ethnic groups: associations with socio-demographic factors. *BMC Public Health, 8*(1), 419. doi: 10.1186/1471-2458-8-419

Sallis, J. F., Bull, F., Guthold, R., Heath, G. W., Inoue, S., Kelly, P., . . . Hallal, P. C. (2016). Progress in physical activity over the Olympic quadrennium. *The Lancet, 388*(10051), 1325-1336. doi: http://dx.doi.org/10.1016/S0140-6736(16)30581-5

Sallis, J. S., & Owen, N. (1998). *Physical Activity and Behavioral Medicine* (Vol. 3). Thousand Oaks, California: Sage Publications.

Samdal, O., Tynjälä, J., Roberts, C., Sallis, J. F., Villberg, J., & Wold, B. (2007). Trends in vigorous physical activity and TV watching of adolescents from 1986 to 2002 in seven European countries. *European Journal of Public Health, 17*(3), 242-248. doi: eurpub/ckl245

Schaefer, M. K., & Salafia, E. H. B. (2014). The connection of teasing by parents, siblings, and peers with girls' body dissatisfaction and boys' drive for muscularity: The role of social comparison as a mediator. *Eating behaviors, 15*(4), 599-608.

Sheridan, D., Coffee, P., & Lavallee, D. (2014). A systematic review of social support in youth sport. *International Review of Sport and Exercise Psychology, 7*(1), 198-228. doi: 10.1080/1750984X.2014.931999

Shiell, A., Hawe, P., & Gold, L. (2008). Complex interventions or complex systems? Implications for health economic evaluation. *BMJ, 336*(7656), 1281-1283. doi: 10.1136/bmj.39569.510521.AD

Siddiqui, F., Lindblad, U., Nilsson, P. M., & Bennet, L. (2019). Effects of a randomized, culturally adapted, lifestyle intervention on mental health among Middle-Eastern immigrants. *European Journal of Public Health, 29*(5), 888-894. doi: 10.1093/eurpub/ckz020

Silverman, M. N., & Deuster, P. A. (2014). Biological mechanisms underlying the role of physical fitness in health and resilience. *Interface Focus, 4*(5). doi: http://dx.doi.org/10.1098/rsfs.2014.0040

Sim, J., & Waterfield, J. (2019). Focus group methodology: some ethical challenges. *Quality & Quantity, 53*(6), 3003-3022. doi: 10.1007/s11135-019-00914-5

Singh, A. S., Saliasi, E., van den Berg, V., Uijtdewilligen, L., de Groot, R. H. M., Jolles, J., . . . 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), 640-647. doi: 10.1136/bjsports-2017-098136

Smith, J., Beauchamp, M., Faulkner, G., Morgan, P., Kennedy, S., & Lubans, D. R. (2018). Intervention effects and mediators of well-being in a school-based physical activity program for adolescents: The 'Resistance Training for Teens' cluster RCT. *Mental Health and Physical Activity, 15*, 88-94. doi: https://doi.org/10.1016/j.mhpa.2018.08.002

Smith, J. J., Eather, N., Morgan, P. J., Plotnikoff, R. C., Faigenbaum, A. D., & Lubans, D. R. (2014). The Health Benefits of Muscular Fitness for Children and Adolescents: A Systematic Review and Meta-Analysis. *Sports Medicine, 44*(9), 1209-1223. doi: 10.1007/s40279-014-0196-4

Smyth, R. L., & Weindling, A. M. (1999). Research in children: ethical and scientific aspects. *The Lancet, 354*, SII21-SII24. doi: https://doi.org/10.1016/S0140-6736(99)90253-2

Spencer, L., & Ritchie, J. (2002). Qualitative data analysis for applied policy research. In A. Bryman & R. G. Burgess (Eds.), *Analyzing qualitative data* (pp. 187-208). London: Routledge.

Stathopoulou, G., Powers, M. B., Berry, A. C., Smits, J. A. J., & Otto, M. W. (2006). Exercise Interventions for Mental Health: A Quantitative and Qualitative Review. *Science and Practice, 13*(2), 179-193. doi: 10.1111/j.1468-2850.2006.00021.x

Steca, P., Abela, J. R. Z., Monzani, D., Greco, A., Hazel, N. A., & Hankin, B. L. (2014). Cognitive Vulnerability to Depressive Symptoms in Children: The Protective Role of Self-efficacy Beliefs in a Multi-Wave Longitudinal Study. *Journal of Abnormal Child Psychology, 42*(1), 137-148. doi: 10.1007/s10802-013-9765-5

Steiger, A. E., Allemand, M., Robins, R. W., & Fend, H. A. (2014). Low and Decreasing Self-Esteem During Adolescence Predict Adult Depression Two Decades Later. *Journal of Personality and Social Psychology, 106*(2), 325–338. doi: 10.1037/a0035133

Steinberg, L. (2014). *Age of opportunity: Lessons from the new science of adolescence*. Boston, MA: Houghton Mifflin Harcourt.

Steinberg, L., & Lerner, R. M. (2004). The scientific study of adolescence: A brief history. *The Journal of Early Adolescence, 24*(1), 45-54. doi: https://doi.org/10.1177/0272431603260879
Steinberg, L., & Morris, A. S. (2001). Adolescent Development. *Annual Review of Psychology*, *52*(1), 83-110. doi: 10.1146/annurev.psych.52.1.83
Sterne, J. A. C., White, I. R., Carlin, J. B., Spratt, M., Royston, P., Kenward, M. G., . . . Carpenter, J. R. (2009). Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *BMJ, 338*, b2393. doi: 10.1136/bmj.b2393
Stoll, L., & Fink, D. (1996). *Changing our schools: Linking school effectiveness and school improvement*: Open University Press.
Streeter, V. M., Milhausen, R. R., & Buchholz, A. C. (2012). Body Image, Body Mass Index, and Body Composition: In Young Adults. *Canadian Journal of Dietetic Practice and Research, 73*(2), 78-83. doi: 10.3148/73.2.2012.78
Stroth, S., Kubesch, S., Dieterle, K., Ruchsow, M., Heim, R., & Kiefer, M. (2009). Physical fitness, but not acute exercise modulates event-related potential indices for executive control in healthy adolescents. *Brain Research, 1269*, 114-124. doi: https://doi.org/10.1016/j.brainres.2009.02.073
Sugai, G., & Horner, R. R. (2006). A promising approach for expanding and sustaining school-wide positive behavior support. *School psychology review, 35*(2), 245-259. doi: https://doi.org/10.1080/02796015.2006.12087989
Sui, X., Laditka, J. N., Church, T. S., Hardin, J. W., Chase, N., Davis, K., & Blair, S. N. (2009). Prospective study of cardiorespiratory fitness and depressive symptoms in women and men. *Journal of Psychiatric Research, 43*(5), 546-552. doi: https://doi.org/10.1016/j.jpsychires.2008.08.002
Sullivan, T. R., White, I. R., Salter, A. B., Ryan, P., & Lee, K. J. (2018). Should multiple imputation be the method of choice for handling missing data in randomized trials? *Statistical Methods in Medical Research, 27*(9), 2610-2626. doi: 10.1177/0962280216683570
Sun, X., Briel, M., Busse, J. W., You, J. J., Akl, E. A., Mejza, F., . . . Guyatt, G. H. (2012). Credibility of claims of subgroup effects

in randomised controlled trials: systematic review. *BMJ, 344*, e1553. doi: 10.1136/bmj.e1553

Tantillo, M., Kesick, C. M., Hynd, G. W., & Dishman, R. K. (2002). The effects of exercise on children with attention-deficit hyperactivity disorder. *Medicine & Science in Sports & Exercise, 34*(2), 203-212. doi: 10.1097/00005768-200202000-00004

Tarokh, L., Saletin, J. M., & Carskadon, M. A. (2016). Sleep in adolescence: Physiology, cognition and mental health. *Neuroscience and biobehavioral reviews, 70*, 182-188. doi: 10.1016/j.neubiorev.2016.08.008

Taylor, R. D., Oberle, E., Durlak, J. A., & Weissberg, R. P. (2017). Promoting Positive Youth Development Through School-Based Social and Emotional Learning Interventions: A Meta-Analysis of Follow-Up Effects. *Child Development, 88*(4), 1156-1171. doi: 10.1111/cdev.12864

Teixeira, P. J., Carraça, E. V., Markland, D., Silva, M. N., & Ryan, R. M. (2012). Exercise, physical activity, and self-determination theory: A systematic review. *International Journal of Behavioral Nutrition and Physical Activity, 9*(1), 78. doi: 10.1186/1479-5868-9-78

Turner, R. J., & Brown, R. L. (2010). Social support and mental health. In T. L. Scheid & T. N. Brown (Eds.), *A handbook for the study of mental health: Social contexts, theories, and systems* (Vol. 2, pp. 200-212). New York: Cambridge University Press.

Van Roy, B., Veenstra, M., & Clench-Aas, J. (2008). Construct validity of the five-factor Strengths and Difficulties Questionnaire (SDQ) in pre-, early, and late adolescence. *The Journal of Child Psychology and Psychiatry, 49*(12), 1304-1312. doi: 10.1111/j.1469-7610.2008.01942.x

Virta, E., Sam, D. L., & Westin, C. (2004). Adolescents with Turkish background in Norway and Sweden: A comparative study of their psychological adaptation. *Scandinavian Journal of Psychology, 45*(1), 15-25. doi: 10.1111/j.1467-9450.2004.00374.x

Vlachopoulos, S. P., Ntoumanis, N., & Smith, A. L. (2010). The basic psychological needs in exercise scale: Translation and evidence for cross-cultural validity. *International Journal of Sport and*

Exercise Psychology, 8(4), 394-412. doi: 10.1080/1612197X.2010.9671960

Voelker, D. K., Reel, J. J., & Greenleaf, C. (2015). Weight status and body image perceptions in adolescents: current perspectives. *Adolescent Health, Medicine and Therapeutics, 6*, 149-158. doi: http://dx.doi.org/10.2147/AHMT.S68344

Voss, M. W., Weng, T. B., Burzynska, A. Z., Wong, C. N., Cooke, G. E., Clark, R., . . . Kramer, A. F. (2016). Fitness, but not physical activity, is related to functional integrity of brain networks associated with aging. *NeuroImage, 131*, 113-125. doi: https://doi.org/10.1016/j.neuroimage.2015.10.044

Wasserstein, R. L., & Lazar, N. A. (2016). The ASA Statement on p-Values: Context, Process, and Purpose. *The American Statistician, 70*(2), 129-133. doi: 10.1080/00031305.2016.1154108

Wasserstein, R. L., Schirm, A. L., & Lazar, N. A. (2019). Moving to a World Beyond "p < 0.05". *The American Statistician, 73*(sup1), 1-19. doi: 10.1080/00031305.2019.1583913

Watson, A., Timperio, A., Brown, H., Best, K., & 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*(114). doi: 10.1186/s12966-017-0569-9

West, B. T. (2009). Analyzing Longitudinal Data With the Linear Mixed Models Procedure in SPSS. *32*(3), 207-228. doi: 10.1177/0163278709338554

Whitehead, M. (2010). *Physical Literacy: Throughout the Lifecourse*. Abingdon, Oxon: Routledge.

Whittemore, R., Chase, S. K., & Mandle, C. L. (2001). Validity in Qualitative Research. *Qualitative Health Research, 11*(4), 522-537. doi: 10.1177/104973201129119299

WHO. (2010). *Global recommendations on physical activity for health*. Switzerland: World Health Organization. Retrieved from https://apps.who.int/iris/bitstream/handle/10665/44399/9789241599979_eng.pdf;jsessionid=DE683C073EA364D35A3D35E8DCCD9B2C?sequence=1

WHO. (2014, August). Mental health: a state of well-being. Retrieved from http://origin.who.int/features/factfiles/mental_health/en/

Wieland, M. L., Tiedje, K., Meiers, S. J., Mohamed, A. A., Formea, C. M., Ridgeway, J. L., . . . Sia, I. G. (2015). Perspectives on Physical Activity Among Immigrants and Refugees to a Small Urban Community in Minnesota. *Journal of Immigrant and Minority Health, 17*(1), 263-275. doi: 10.1007/s10903-013-9917-2

Wijndaele, K., Matton, L., Duvigneaud, N., Lefevre, J., De Bourdeaudhuij, I., Duquet, W., . . . Philippaerts, R. M. (2007). Association between leisure time physical activity and stress, social support and coping: A cluster-analytical approach. *Psychology of Sport and Exercise, 8*(4), 425-440. doi: https://doi.org/10.1016/j.psychsport.2006.08.001

Wiles, N. J., Jones, G. T., Haase, A. M., Lawlor, D. A., Macfarlane, G. J., & Lewis, G. (2008). Physical activity and emotional problems amongst adolescents. *Social Psychiatry and Psychiatric Epidemiology, 43*(10), 765. doi: 10.1007/s00127-008-0362-9

Wills, T. A., Isasi, C. R., Mendoza, D., & Ainette, M. G. (2007). Self-Control Constructs Related to Measures of Dietary Intake and Physical Activity in Adolescents. *Journal of Adolescent Health, 41*(6), 551-558. doi: https://doi.org/10.1016/j.jadohealth.2007.06.013

Woodward, L. J., & Fergusson, D. M. (1999). Childhood Peer Relationship Problems and Psychosocial Adjustment in Late Adolescence. *Journal of Abnormal Child Psychology, 27*(1), 87-104. doi: 10.1023/a:1022618608802

Xie, H., Swift, D. J., Cairns, B. D., & Cairns, R. B. (2002). Aggressive Behaviors in Social Interaction and Developmental Adaptation: A Narrative Analysis of Interpersonal Conflicts During Early Adolescence. *Social Development, 11*(2), 205-224. doi: 10.1111/1467-9507.00195

Yeatts, P. E., Martin, S. B., & Petrie, T. A. (2017). Physical fitness as a moderator of neuroticism and depression in adolescent boys and girls. *Personality and Individual Differences, 114*, 30-35. doi: 10.1016/j.paid.2017.03.040

Youthinmind. (2018). *Information for researchers and professionals about the Strengths & Difficulties Questionnaires.* Retrieved from http://www.sdqinfo.com/

Zhang, H., Li, L., Wu, M., Chen, Z., Hu, X., Chen, Y., . . . Gong, Q. (2016). Brain gray matter alterations in first episodes of depression: A meta-analysis of whole-brain studies. *Neuroscience & Biobehavioral Reviews, 60*, 43-50. doi: https://doi.org/10.1016/j.neubiorev.2015.10.011

Part 2

The Papers

Paper 1

Åvitsland et al. BMC Public Health (2020) 20:776
https://doi.org/10.1186/s12889-020-08936-7

BMC Public Health

RESEARCH ARTICLE Open Access

The association between physical fitness and mental health in Norwegian adolescents

Andreas Åvitsland[1*], Eva Leibinger[1], Tommy Haugen[2], Øystein Lerum[3], Runar B. Solberg[4], Elin Kolle[4] and Sindre M. Dyrstad[1,5]

Abstract

Background: Studies indicate that health-related components of physical fitness are associated with mental health outcomes. However, research is scarce concerning this relationship in young adolescents in general and non-existent in Norwegian populations specifically. The aim of the study was to examine whether body composition, muscular strength and cardiorespiratory fitness were associated with self-reported mental health in Norwegian adolescents.

Methods: Adolescents from four regions of Norway ($n = 1486$; mean age = 13.9; girls = 50.6%) participated. Self-reported mental health (psychological difficulties) was measured by completing the Strengths and Difficulties Questionnaire. Cardiorespiratory fitness was assessed with an intermittent running test; muscular strength was assessed by measuring handgrip strength, standing broad jump and sit-ups; and body composition was assessed by calculating body mass index from weight and height. Linear mixed effects models were conducted to assess the associations between the health-related components of physical fitness and psychological difficulties. School clusters were included as random effects and all models were controlled for sex, socioeconomic status and birthplace (domestic or foreign).

Results: Body composition was not associated with psychological difficulties. Muscular strength was independently associated with psychological difficulties, but when all independent variables were entered in the fully adjusted model, only cardiorespiratory fitness was associated with psychological difficulties.

Conclusions: There was a small but significant inverse association between cardiorespiratory fitness and levels of psychological difficulties in Norwegian adolescents. The results suggest that muscular strength is not associated with psychological difficulties in adolescents, when controlling for cardiorespiratory fitness. Future research should focus on the prospective association between physical fitness components and mental health outcomes in adolescents.

Trial registration: The study is registered in ClinicalTrials.gov ID nr: NCT03817047. Retrospectively registered January 25, 2019.

Keywords: Physical fitness, Mental health, Strengths and difficulties, Norway, Adolescence, Cross-sectional

* Correspondence: Andreas.avitsland@uis.no
[1]Department of Education and Sport Science, University of Stavanger, 4036 Stavanger, Norway
Full list of author information is available at the end of the article

Background

"Mental health is defined as a state of well-being in which every individual realizes his or her own potential, can cope with the normal stresses of life, can work productively and fruitfully, and is able to make a contribution to her or his community" [1]. Mental health problems affect between 10 and 20% of the global child and adolescent population [2] and approximately 18% of adults will have experienced a form of mental disorder within a one-year period [3]. Depression is the most prevalent mental disorder, accounting for 41% of all disability-adjusted life years caused by mental and substance use disorders [4]. Adolescent mental health problems have increased during recent decades in middle- and high-income countries [5]. Although part of the increase may be attributed to more awareness, help-seeking, and a lower threshold for treatment, a real prevalence increase for mental health problems has likely occurred. In Norway, the percentage of lower secondary school girls reporting depressive symptoms has increased from 16% in 2011 to 20% in 2016 [6]. Furthermore, at the end of upper secondary school, 12% of boys and 29% of girls display high levels of depressive symptoms. Onset of depression during adolescence is associated with poor general health, increased work impairment and higher utilization of health care services at age 20 [7], which has contributed to mental disorders being the costliest conditions in Norway [8]. Considering these detrimental effects of adolescent mental health problems, it is important to find effective methods of prevention, or methods to improve adolescent mental health.

Physical activity, physical fitness and potential mechanisms

Evidence suggests physical activity is a protective factor against mental health problems such as depression [9]. Moreover, physical activity can positively affect a range of other mental health outcomes, such as mood, stress, cognitive functioning [10], and self-worth [11]. Although there is no clear consensus, many mechanisms have been hypothesized to explain the relationship between physical activity and mental health. Lubans et al. [12] elucidated three of these mechanisms. First, the neurobiological mechanism proposes that physical activity alters structural and functional compositions of the brain. Second, the psychosocial mechanism proposes that physical activity can provide social interaction, physical mastery, independence and improved appearance self-perception. Lastly, the behavioral mechanism proposes that changes in behavior, such as sleep and coping skills, mediates how physical activity affects mental health outcomes.

While our physical activity level fluctuates from week to week, physical fitness, although somewhat influenced by genes [13], represents the type, frequency, intensity and duration of physical activity that has occurred over time [14]. Physical fitness may therefore provide a more stable measure of habitual physical activity levels. Physical fitness can be divided into health-related components, such as cardiorespiratory fitness (CRF), muscular strength, and body composition [15]. Given the relationship between physical activity and physical fitness, the mechanisms proposed to explain the relationship between physical activity and mental health might also apply for physical fitness and mental health [16]. However, the aforementioned components of physical fitness may also be associated with mechanisms influencing mental health outcomes, independent from physical activity [17]. High CRF, as a result of vigorous aerobic physical activity [18] can affect neurobiological processes and inhibit inflammation [16]. Body composition affects body image [19], which in turn depends on cultural norms [20]. Therefore, body composition may affect mental health outcomes through sociocultural or psychosocial mechanisms, in addition to the possible biological mechanisms that are associated with obesity [21]. Muscular strength may also depend on cultural norms [22], thereby possibly affecting mental health through similar mechanisms as body composition. Additionally, muscular strength may affect neurobiological processes differently than CRF, however, this is unclear [23].

Mental health and physical fitness in adolescents

Many studies have been conducted with adult populations regarding associations between mental health outcomes and physical fitness. For instance, reviews show that lower levels of mental disorders have been associated with higher CRF [24], muscular strength [25] and healthier body composition [26]. Similar studies examining adolescent populations, however, are scarce. Ruggero et al. [27], showed that CRF was inversely associated with depression in 12- and 13-year-old girls ($r = -.31$) and boys ($r = -.39$). Another study found that adolescent girls with low CRF exhibited 31% higher levels of depression, compared to girls with high CRF [28]. The same study also showed that boys and girls categorized as having a fit body composition exhibited 12 and 25% higher body satisfaction, respectively, compared to students categorized as having an unfit body composition. Regarding muscular strength, Lubans and Cliff [29] found an association with self-worth in boys but not girls, and a review by Smith et al. [30] showed a strong association with self-esteem in adolescent populations.

Aim

A small amount of evidence regarding adolescents indicates a relationship between the components of health-related physical fitness and mental health outcomes.

However, to the best of our knowledge, only the study by Yeatts, Martin and Petrie [31] has measured the three components CRF, muscular strength and body composition in association with a mental health outcome in adolescents. Thus, it is unclear whether one component is more important than others. Regarding Norwegian adolescents specifically, only one study has examined a mental health outcome in association with physical fitness [32]. Therefore, the aim of the present paper was to investigate the relationship between health-related components of physical fitness and mental health in Norwegian adolescents.

Methods

Design and participants

The present study used cross-sectional data from the baseline tests of the School in Motion project [33]. This was a multicenter study, involving four geographically separate regional test centers in Norway. Out of 103 invited lower secondary schools, 29 schools agreed to participate. Only eighth grade students (13–14-year-olds) were invited to participate in the study ($n = 2733$). Informed parental consent was obtained from 76% of the invited students ($n = 2084$). Not all students had valid measures on all variables and Fig. 1 shows an overview of the participant flow. The participants were tested in the spring of 2017, during school time, at their respective schools. All test personnel received the same training beforehand to make sure there were no discrepancies in how the tests were carried out. All test procedures were approved by the Norwegian Centre for Research Data (project number 49094), and the project is in accordance with the Declaration of Helsinki for experiments involving humans.

Measurements

Body composition

Participants' weight without shoes was measured by digital scale (Seca 899, Hamburg, Germany) and all measurements were recorded to the closest 0.1 kg. Their clothes were noted, and their weight adjusted in the following analysis: 1 kg was subtracted for pants and/or sweater, 0.5 kg was subtracted for shorts/tights and t-shirt. Height was measured by portable stadiometer

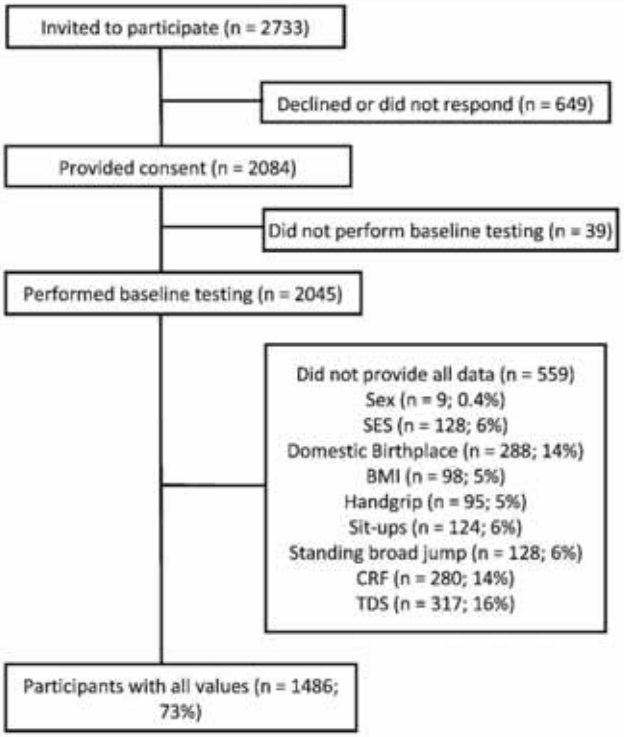


Fig. 1 Flow chart of recruitment and participation with an overview of missing values. SES = socioeconomic status, BMI = body mass index, CRF = cardiorespiratory fitness, TDS = total difficulties score

(Seca 123, Hamburg, Germany) and was recorded to the closest mm. The values were used to calculate individual body mass index (BMI) scores (kg/m^2). None of the measurements were disclosed to the participants.

Muscular strength

Sit-ups (n/30 s), standing broad jump (best of two attempts) and handgrip test (best of two attempts), as described in the EUROFIT test battery [34] were used to measure muscular strength. Participants performed sit-ups with their knees in a 90-degree angle and their fingers locked behind their head, and their feet held to the floor by test personnel. To get a valid count, the participants had to touch their knees with their elbows, going up, and touch the floor with their shoulders, going down. Participants performed standing broad jump by jumping as far as they could from a stand still position, and the distance was recorded from the heel closest to the starting point. Measurements were recorded to the closest cm. The handgrip strength test was executed with the participants' dominant hand, as they held their arm down alongside their body, gripping a Baseline dynamometer (Baseline® Hydraulic Hand Dynamometer, Elmsford, NY, USA) as hard as they could for 3 s. Measurements were recorded to the closest kg.

Cardiorespiratory fitness

CRF was assessed by a 10-min intermittent running test [35]. The test was performed by the participants running between two marked lines, 16 m apart, inside a gymnasium. They were required to touch the floor behind the line with one hand before turning and running back. The participants ran for 15 s, then paused for 15 s on the test leader's whistle and this procedure was repeated for 10 min. According to test protocol, the intended distance between the lines is 20 m; however, limited space in many school gymnasiums compelled us to set a new standard distance at 16 m. Because of this, we could not estimate maximum oxygen uptake from the test results, therefore, we use running distance in meters (m) as an indirect measurement unit of CRF when describing our results.

Mental health

To measure mental health, the participants completed a Norwegian language version of the Strengths and Difficulties Questionnaire [SDQ [36];]. The questionnaire consists of 25 items divided into five subscales. The five subscales cover emotional symptoms, conduct problems, hyperactivity, peer relationships and prosocial behavior. The questionnaire contains statements such as "I worry a lot"," I am easily distracted, I find it difficult to concentrate" and" Other people my age generally like me". Participants reply to the statements on a three-point Likert scale:" not true"," somewhat true" and" certainly true". Each subscale scores from 0 to 10. Except for the prosocial subscale, a higher score signifies a higher degree of difficulties. A high score on the prosocial subscale signifies social strengths. The scores from all subscales except the prosocial are summed to create the total difficulties score (TDS). TDS scores from 0 to 40 and is a dimensional measure of mental health for children and adolescents, which means that on a population level, there is a detectable reduction in psychopathology for each point-reduction on the scale [37]. It therefore represents an indication of the general mental health state in the measured population, but in the continuation of the paper, we will refer to the outcome as either TDS or psychological difficulties. The psychometric properties of the SDQ have been validated in several countries [38–40], including Norway [41].

Covariates

Other variables associated with mental health are sex [42], domestic or foreign birthplace [43], and socioeconomic status [SES [44];]. The participants' sex was noted by test-personnel, and birthplace ("Were you born in Norway") was assessed in the questionnaire. Parents' education level was included as a measure of SES [45].

Statistical analysis

Data management

Data were managed and analyzed in IBM SPSS Statistics 25 (IBM, Armonk, New York, USA). SDQ data were scored according to the syntax provided by the SDQ information web page [46]. The syntax summed the scores from each of the four subscales needed to create the TDS variable. Cronbach's alpha was employed to assess the internal consistency of TDS and the result was .62.

We created z scores stratified for sex and BMI quartiles for handgrip strength, standing broad jump and sit-ups. The z scores were used to create one composite mean z score for muscular strength. SES was analyzed by including only the parent with the highest education level. Next, parents' education level was categorized as either" lower secondary school or less"," upper secondary school", "less than four years university education" and "four years or more university education".

Out of 2045 participants, 27% (n = 559; girls = 38.2%) had at least one missing value. A new grouping variable was created to analyze differences between participants with all values (n = 1486) and participants with missing values (n = 559). The following primary analyses were carried out on the complete-case group only, while extensive missing value analyses were conducted to examine if they influenced the primary results.

Complete-case primary analyses

Descriptive statistics were calculated and are presented as means and standard deviations (SD). Seven linear mixed effect models with TDS as the dependent variable were conducted. In models one to six, we assessed the separate associations between TDS and the muscular strength variables and the health-related fitness components. In the seventh model, the fitness components controlled for each other. All models controlled for the covariates (sex, domestic birthplace and SES). We report estimates (unstandardized coefficients) and their 95% confidence intervals (95% CI). Estimates reflect the change in TDS as a result of one unit of measurement increase in the independent variables. Initial linear mixed effect modelling showed no statistically significant interaction effects between sex and the physical fitness variables, using TDS as the dependent variable. To account for possible effects of clustering of observations within schools, school site was included as a random effect in all models. A p value < .05 indicated statistical significance.

Missing value analyses

To assess whether missing values were missing completely at random (MCAR), Little's MCAR test was used. The analysis did not support MCAR (104.331, DF = 24, $p < .001$). Pattern analysis (not shown) indicated that the data were likely missing at random (MAR). A possible explanation for the missing values is that we never forced the participants to complete the tests, which may have caused some participants to opt out. For instance, many stated that they did not want to run the CRF test. Moreover, the SDQ was one of many components in a large and extensive questionnaire. The missing data from the SDQ may be a consequence of the size and duration of the extended questionnaire, which may have caused many to quit before completion. However, this is unclear and there may be other reasons unknown to us.

One-way ANOVA was used to assess differences between the complete-case group and the missing-values group. Pearson's correlation analysis was used on the fitness variables and TDS, for the purpose of examining if associations were similar in both groups. Multicenter studies are vulnerable to differences in missingness between test centers [47], and this was examined using frequency statistics. As our final action in handling the missing values, we employed multiple imputation [48, 49]. Five imputations were generated from relevant variables, using the automatic procedure with 10 iterations, with the assumption that data were missing at random. A linear mixed effects model was conducted on the imputed dataset, with TDS as the dependent variable, and all health-related components of physical fitness variables and covariates entered as independent variables. The imputed dataset results are presented, in addition to the complete-case results, as recommended by Manly and Wells [50] and Sterne et al. [51].

Results

Descriptives and group comparisons

Descriptive results and group differences between the complete-case group and missing-values group are presented in Table 1. Compared to the complete-case group, the missing-values group had 1.5% higher mean BMI ($p = .047$), performed 5.9% worse on the handgrip test ($p < .001$), 2.3% worse on the standing broad jump test ($p = .005$) and had 2.9% higher CRF ($p < .001$). The missing-values group scored 15.5% higher for TDS ($p < .001$).

Linear mixed effects models

Results from the linear mixed effects models are summarized in Table 2. Model 1 indicated no association between BMI and TDS. Model 2 indicated an inverse association between sit-ups and TDS (b = −.088; 95% CI = −.156 to −.020; $p = .011$). Model 3 indicated no association between standing broad jump and TDS. Model 4 indicated no association between handgrip strength and TDS. Model 5 indicated an inverse association between muscular strength and TDS (b = −.458; 95% CI = −.810 to −.109; $p = .010$) Model 6 indicated an inverse association between CRF and TDS (b = −.006; 95% CI = −.009 to −.003; $p < .001$). The fully adjusted model (7) with all independent variables and covariates entered simultaneously, revealed no association between muscular strength and TDS, while the association between CRF and TDS remained almost identical as in Model 6 (b = −.006; 95% CI = −.009 to −.002; $p = .001$). The estimates produced by the fully adjusted model suggest that every 100 m distance increase in the CRF test is associated with 1.5% (0.6 points) lower TDS.

Results from missing value analyses

Among the 559 excluded participants, 12.2% ($n = 68$) had completed the SDQ and performed the CRF test. Correlation analysis indicated a non-significant, inverse correlation between CRF and TDS ($r = -.13$; $p = .139$), a similar relationship as in the complete-case group. Other correlations were also similar (data not shown), indicating a small likelihood of systematic differences between the complete-case group and missing-values group.

There was a difference in missingness between test centers. One test center represented 10.7% of the participants in the complete-case group and 30.4% in the missing-values group, which means that more than half of all participants from this region were excluded due to missing values. This was caused by low completion of the SDQ, with only 13.5% of these participants completing the SDQ in the missing-values group. The

Table 1 Participant mean characteristics from the complete-case group and the missing-values group. Presented as means with standard deviations (SD)

	Total difficulties score (0–40)	Body mass index	Sit-ups (n/30 s)	Standing broad jump (cm)	Handgrip strength (kg)	Cardiorespiratory fitness (meters)
All values present						
Yes (n = 1486)	10.3 (5.1)	19.8 (3.1)	19 (4)	172.2 (25.9)	30.5 (7.0)	906 (98)
No (n = 242–559)	11.9 (5.7)[a]	20.2 (3.5)[a]	19 (4)	168.2 (27.8)[a]	28.7 (7.4)[a]	932 (115)[a]

Note. [a]Difference between groups is significant at *p* level < .05

corresponding completion rates from other test centers were 51.2–67.6%.

Finally, the pooled dataset from the multiple imputation (n = 2045) was analyzed with a fully adjusted linear mixed effects model, in the same way as the complete-case dataset. The association between CRF and TDS was close to identical to the complete-case results (b = −.006; 95% CI: −.010 to −.002; p = .006). Thus, all missing-value analyses indicated that the results likely would have been unchanged with all values present.

Discussion

The main findings of the present study were that higher CRF was significantly associated with lower psychological difficulties in Norwegian adolescents while controlling for muscular strength, body composition, socioeconomic status, school clustering, sex and domestic/foreign birthplace. The results indicated that psychological difficulties were not associated with muscular strength or body composition, when controlling for the aforementioned variables.

Muscular strength, body mass index and metal health

A significant association between muscular strength and TDS was initially found. However, when controlling for CRF there was no association between these variables. A possible explanation is that the participants with high CRF were also likely to have a relatively high muscular strength [52]. However, based on the fully adjusted model, it can be postulated that participants with high muscular strength did not necessarily have high CRF. A possible interpretation is that muscular strength in adolescents is generally a natural consequence of the individual's CRF level, which may represent the true association with psychological difficulties.

The present findings support previous studies that have suggested CRF to be the only health-related aspect of fitness associated with mental health outcomes such as quality of life [53], depression [31, 54] and well-being [55]. Many studies that have found associations between muscular strength and mental health outcomes in adolescents did not measure CRF [29, 56]. The present findings did not show an association between BMI and psychological difficulties, independent from controlling for CRF. This is congruent with the review by Luppino et al. [21], who found an association between overweight and depression in adults, but not in individuals younger than 20 years. This indicates a different relationship between age groups; however, it is important to point out that none of the reviewed studies controlled for CRF.

Table 2 Linear mixed effect models with Total difficulties score as the dependent variable

Model	Independent variable	B (95% CI)	p
1	Body mass index	.058 (−.027 to .142)	.184
2	Sit-ups	−.088 (−.156 to −.020)	.011
3	Standing broad jump	−.008 (−.018 to .003)	.147
4	Handgrip strength	−.024 (−.062 to .013)	.206
5	Muscular strength z score	−.458 (−.810 to −.109)	.010
6	Cardiorespiratory fitness	−.006 (−.009 to −.003)	<.001
7 – Fully adjusted model	Intercept	18.8 (14.6 to 23.0)	<.001
	Body mass index	.003 (−.086 to .093)	.941
	Muscular strength z score	−.033 (−.450 to .381)	.873
	Cardiorespiratory fitness	−.006 (−.009 to −.002)	.001

Note. 29 school clusters were included as random effects in all models; all models were controlled for sex, socioeconomic status, and birthplace. *B* Unstandardized regressioncoefficient, *CI* Confidence interval, *p* Significance level

Although it is possible that muscular strength and BMI are associated with mental health outcomes, studies that do not also measure CRF lack important information. Had we not controlled for all fitness variables in the present study, we would have erroneously concluded that muscular strength was associated with TDS. Opposing findings by Kettunen et al. [57] showed muscular strength to be more important than CRF to reduce stress in adults. However, this study categorized continuous variables and employed ANOVAs, which has been strongly advised against by Altman and Royston [58] and might have produced biased results. Additionally, associations between physical fitness and mental health might be different in adult and adolescent populations. For instance, many experimental studies have found effects of strength training on mental health in older adults [23]. Positive effects in older adults are not surprising, considering how strength training can reverse muscle atrophy and improve the daily functioning of older people [59]. In adolescents however, muscular strength is mainly associated with appearance-related mental health outcomes, such as self-perception, perceived physical appearance or physical self-worth [30]. Future studies of associations between health-related physical fitness and mental health should include different mental health outcomes, to gain a better understanding of whether specific components of fitness are associated with specific outcomes of mental health.

Cardiorespiratory fitness and mental health

Although a causal direction between CRF and psychological difficulties cannot be established from cross-sectional findings, recent evidence has indicated a one-directional causal relationship for physical activity as a protective factor against depression among adults [60]. High-intensity exercise is an important factor for high CRF [18], hence results from the present study support a hypothesis suggesting that high-intensity exercise might be more favorable for mental health than low-intensity exercise. This is in accordance with the study by Parfitt, Pavey and Rowlands [61], who found high-intensity exercise to be more favorable for mental health than light-intensity exercise, in a population of children. Furthermore, the meta-analysis by Ahn and Fedewa [62] found high-intensity exercise RCT interventions to have the most effect on children's mental health. On the other hand, Helgadóttir et al. [63] concluded that low-intensity exercise was more effective on depression treatment than high-intensity exercise in an adult population. The low-intensity group exercised with yoga and this type of exercise may have a distinct relationship with mental health. However, the results should be treated with caution, because the intervention had low adherence and did not mention how this differed between exercise groups. Additionally, 12 months after the intervention, there were no significant differences between the low- and vigorous-exercise groups. The study by Helgadóttir et al. [63] is incongruent with the previously mentioned studies, as well as what Bailey et al. [64] suggested to treat depression in adolescents: "… aerobic-based activity of moderate-to-vigorous intensity." It is also possible that intensity might not even be especially crucial, as long as CRF is improved. Shepherd et al. [65] prescribed high-intensity interval training and moderate-intensity continuous training in two groups of inactive adults and both groups experienced increased CRF and improved mental health. Few studies have examined the causal relationship between increased CRF and improved mental health outcomes, but a recent longitudinal study by Rahman et al. [66] showed that improved CRF predicted at least a 50% reduction in depression scores for adults. Ruggero et al. [27] found that high CRF at baseline was associated with lower levels of depression a year later in adolescent girls and suggested that CRF might mediate the effect physical activity has on depression. This was supported by Eddolls et al. [67] who concluded that CRF mediated the relationship between vigorous physical activity and mental health in adolescents, thus suggesting that physical activity interventions to treat depression may only be effective if they improve CRF.

Research on the potential explanatory mechanisms between muscular strength and mental health is scarce [23]. There are, however, mechanisms that might explain the association between CRF and psychological difficulties. One example is the endocannabinoid system, which mediates high-intensity aerobic exercise effect on depression [68]. Psychosocial mechanisms may also have had a mediating role in the present results: CRF is associated with team sports like football, handball and basketball [69], which are important arenas for social relationships and can provide opportunities to improve self-esteem and body satisfaction [12]. The topic of explanatory mechanisms between physical fitness and mental health outcomes requires more research, especially on adolescent populations, in order to fully understand the relationship between the relevant variables. Additionally, future studies need to examine how exercise at different intensities affects different mental health outcomes, and whether the results are influenced by increases in CRF. Such knowledge can be useful in efforts to prevent or treat mental disorders.

Strengths and limitations

Strengths of the present study include the large sample size from separate geographical regions, the use of three objectively measured health-related components of physical fitness, and the control of relevant covariates.

The main limitation of the present study was a large number of missing values; however, the extensive missing value analyses indicated that the main results most likely were unaffected by the dropouts. Reducing the

length of the CRF-test distance is also a limiting factor, as it makes our CRF-results incomparable to results from other studies. However, this limitation did not affect the main results since running distance in meters was used as the measurement unit, and not estimated maximum oxygen uptake. Moreover, we did not measure maturity status, which may act as a confounder for the associations in the main results.

The present internal consistency results pertaining to SDQ were quite low and are similar to Italian [70], Finnish [71] and Dutch [40] results. Internal consistency results from English speaking populations [72, 73] are usually higher, which suggests that statements are better understood by native English speakers, while non-native English speakers may misinterpret the statements somewhat. Age is also a factor, as the internal consistency is lower for younger adolescents, such as the present population, compared to older adolescents as examined in studies by Bøe et al. [74] and Sagatun et al. [75]. Finally, the cross-sectional nature of the study limits the ability to make any causal inference.

Conclusion

The main findings from the present study was that higher cardiorespiratory fitness was significantly associated with lower levels of psychological difficulties in adolescents. Body composition was not associated with psychological difficulties. Muscular strength was separately associated with psychological difficulties but not when controlling for cardiorespiratory fitness. This indicates that strength training or focus on weight reduction may be ineffective in efforts to prevent or treat mental health problems in adolescents. Future research in this area should examine the prospective associations between physical fitness components and mental health outcomes and explore potential reasons why cardiorespiratory fitness seems to be more important for adolescent mental health outcomes than muscular strength and body composition.

Abbreviations

CRF: Cardiorespiratory fitness; BMI: Body mass index; SDQ: Strengths and Difficulties Questionnaire; TDS: Total Difficulties Score; SES: Socioeconomic status; MCAR: Missing completely at random; MAR: Missing at random

Acknowledgements

None.

Authors' contributions

EK, TH, RBS and SMD were involved in the conception and design of the School in Motion study. AÅ, EL, ØL and SMD participated in the conception and design of this paper. All authors participated in the collection and analyses of data, writing of the paper and approved the final version.

Funding

The study was funded by grants from the Norwegian Directorate for Education and Training. The funding body has not influenced the study design, collection, analysis and interpretation of data, or how the manuscript was written.

Availability of data and materials

The datasets generated and/or analyzed during the current study are not publicly available as publications are planned but are available from the corresponding author on reasonable request.

Ethics approval and consent to participate

The project was reviewed by the Regional Committee for Medical and Health Research Ethics (REK) in Norway, who according to the Act on medical and health research (the Health Research Act 2008) concluded that the study did not require full review by REK. The study was approved by the Norwegian Centre for Research Data. Written informed consent from the participants and their parents or caretakers was obtained prior to the data collection.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

[1]Department of Education and Sport Science, University of Stavanger, 4036 Stavanger, Norway. [2]Department of Sport Science and Physical Education, University of Agder, 4604 Kristiansand, Norway. [3]Department of Sport, Food and Natural Sciences, Western Norway University of Applied Sciences, 6851 Sogndal, Norway. [4]Department of Sports Medicine, Norwegian School of Sports Sciences, 0806 Oslo, Norway. [5]Department of Public Health, University of Stavanger, 4036 Stavanger, Norway.

Received: 14 January 2020 Accepted: 17 May 2020
Published online: 24 May 2020

References

1. WHO. Mental health: a state of well-being: World Health Organization; 2014 [updated 2014 08-15 15:25:38. Available from: http://www.who.int/features/factfiles/mental_health/en/.
2. Kieling C, Baker-Henningham H, Belfer M, Conti G, Ertem I, Omigbodun O, et al. Child and adolescent mental health worldwide: evidence for action. Lancet. 2011;378:1515–25.
3. Steel Z, Marnane C, Iranpour C, Chey T, Jackson J, Patel V, et al. The global prevalence of common mental disorders: a systematic review and meta-analysis 1980-2013. Int J Epidemiol. 2014;43(2):476–93.
4. Whiteford HA, Degenhardt L, Rehm J, Baxter AJ, Ferrari AJ, Erskine HE, et al. Global burden of disease attributable to mental and substance use disorders: findings from the global burden of disease study 2010. Lancet. 2013;382(9904):1575–86.
5. Collishaw S. Annual research review: secular trends in child and adolescent mental health. J Child Psychol Psychiatry. 2015;56(3):370–93.
6. Ungdata BA. Nasjonale resultater 2017. NOVA: Oslo; 2017.
7. Keenan-Miller D, Hammen CL, Brennan PA. Health outcomes related to early adolescent depression. J Adolesc Health. 2007;41(3):256–62.
8. Kinge JM, Sælensminde K, Dieleman J, Vollset SE, Of N. Economic losses and burden of disease by medical conditions in Norway. Health Policy. 2017;121(6):691–8.
9. Mammen G, Faulkner G. Physical activity and the prevention of depression: a systematic review of prospective studies. Am J Prev Med. 2013;45(5):649–57.
10. Basso JC, Suzuki WA. The effects of acute exercise on mood, cognition, neurophysiology, and neurochemical pathways: a review. Brain Plasticity. 2017;2(2):127–52.
11. Mingli L, Lang W, Qingsen M. How does physical activity intervention improve self- esteem and self- concept in children and adolescents? Evidence from a meta-analysis. PLoS One. 2015;10(8):e0134804.
12. Lubans DR, Richards J, Hillman C, Faulkner G, Beauchamp M, Nilsson M, et al. Physical activity for cognitive and mental health in youth: a systematic review of mechanisms. Pediatrics. 2016;138(3):e20161642.
13. Fisher A, Smith L, van Jaarsveld CHM, Sawyer A, Wardle J. Are children's activity levels determined by their genes or environment? A systematic review of twin studies. Prev Med Rep. 2015;2:548–53.

14. Blair SN, Cheng Y, Holder JS. Is physical activity or physical fitness more important in defining health benefits? Med Sci Sports Exerc. 2001;33(6 Suppl):379–99.
15. Caspersen CJ, Powell KE, Christenson GM. Physical activity, exercise and physical fitness: definitions and distinctions for health-related research. Public Health Rep. 1985;100(2):126–30.
16. Silverman MN, Deuster PA. Biological mechanisms underlying the role of physical fitness in health and resilience. Interface Focus. 2014;4(5):20140040.
17. Voss MW, Weng TB, Burzynska AZ, Wong CN, Cooke GE, Clark R, et al. Fitness, but not physical activity, is related to functional integrity of brain networks associated with aging. NeuroImage. 2016;131:113–25.
18. Silva G, Andersen LB, Aires L, Mota J, Oliveira J, Ribeiro JC. Associations between sports participation, levels of moderate to vigorous physical activity and cardiorespiratory fitness in children and adolescents. J Sports Sci. 2013;31(12):1359–67.
19. Streeter VM, Mihausen RR, Buchholz AC. Body image, body mass index, and body composition in young adults. Can J Diet Pract Res. 2012;73(2):78–83.
20. Grabe S, Ward LM, Hyde JS. The role of the media in body image concerns among women: a meta-analysis of experimental and correlational studies. Psychol Bull. 2008;134(3):460–76.
21. Luppino FS, de Wit LM, Bouvy PF, Stijnen T, Cuijpers P, Penninx BWJH, et al. Overweight, obesity, and depression: a systematic review and meta-analysis of longitudinal studies. Arch Gen Psychiatry. 2010;67(3):220–9.
22. Rodgers RF, Ganchou C, Franko DL, Chabrol H. Drive for muscularity and disordered eating among French adolescent boys: a sociocultural model. Body Image. 2012;9(3):318–23.
23. O'Connor PJ, Herring MP, Caravalho A. Mental health benefits of strength training in adults. Am J Lifestyle Med. 2010;4(5):377–96.
24. Schuch FB, Vancampfort D, Sui X, Rosenbaum S, Firth J, Richards J, Ward PB, Stubbs B. Are lower levels of cardiorespiratory fitness associated with incident depression? A systematic Review of Prospective Cohort Studies. Prev Med. 2016;93:159–65.
25. Volaklis K, Mamadjanov T, Meisinger C, Linseisen J. Association between muscular strength and depressive symptoms. Wien Klin Wochenschr. 2019;131(11):255–64.
26. Xu Q, Anderson D, Lurie-Beck J. The relationship between abdominal obesity and depression in the general population: A systematic review and meta-analysis. Obes Res Clin Pract. 2011;5(4):e267–78.
27. Ruggero CJ, Petrie T, Sheinbein S, Greenleaf C, Martin S. Cardiorespiratory fitness may help in protecting against depression among middle school adolescents. J Adolesc Health. 2015;57(1):60–5.
28. Greenleaf C, Petrie T, Martin SB. Psychosocial variables associated with body composition and cardiorespiratory fitness in middle school students. Res Q Exerc Sport. 2010;81:65–74.
29. Lubans DR, Cliff DP. Muscular fitness, body composition and physical self-perception in adolescents. J Sci Med Sport. 2011;14(3):216–21.
30. Smith JJ, Eather N, Morgan PJ, Plotnikoff RC, Faigenbaum AD, Lubans DR. The health benefits of muscular fitness for children and adolescents: a systematic review and meta-analysis. J Sports Medicine. 2014;44(9):1209–23.
31. Yeatts PE, Martin SB, Petrie TA. Physical fitness as a moderator of neuroticism and depression in adolescent boys and girls. Personal Individ Differ. 2017;114:30–5.
32. Haugen T, Ommundsen Y, Seiler S. The relationship between physical activity and physical self-esteem in adolescents: the role of physical fitness indices. Pediatr Exerc Sci. 2013;25(1):138–53.
33. Kolle E, Solberg RB, Grydeland M, Säfvenbom R, Andersen SA, Ekelund U, Berntsen S, Resaland GK, Dyrstad SM, Steene-Johannessen J. Evaluating the effect of increased physical activity in lower secondary school adolescents: The study protocol for the School in Motion (SciM) study – a cluster-randomized controlled trial. Manuscript submitted for publication; 2020.
34. Council of Europe. Handbook of the Eurofit tests of Physical Fitness. Roma Italian National. 1988.
35. Andersen L, Andersen TE, Andersen E, Anderssen S. An intermittent running test to estimate maximal oxygen uptake: the Andersen test. J Sports Med Phys Fitness. 2008;48(4):434–7.
36. Goodman R. The strengths and difficulties questionnaire: a research note. J Child Psychol Psychiatry. 1997;38(5):581–6.
37. Goodman A, Goodman R. Strengths and difficulties questionnaire as a dimensional measure of child mental health. J Am Acad Child Adolesc Psychiatry. 2009;48(4):400–3.
38. Rothenberger A, Becker A, Erhart M, Wille N, Ravens-Sieberer U. Psychometric properties of the parent strengths and difficulties questionnaire in the general population of German children and adolescents: results of the BELLA study. Eur Child Adolesc Psychiatry. 2008; 17(1):99.
39. Goodman R. Psychometric properties of the strengths and difficulties questionnaire. J Am Acad Child Adolesc Psychiatry. 2001;40(11):1337–45.
40. Muris P, Meesters C, Eijkelenboom A, Vincken M. The self-report version of the strengths and difficulties questionnaire: its psychometric properties in 8- to 13-year-old non-clinical children. Br J Clin Psychol. 2004;43:437–48.
41. Van Roy B, Veenstra M, Clench-Aas J. Construct validity of the five-factor strengths and difficulties questionnaire (SDQ) in pre-, early, and late adolescence. J Child Psychol psychiatry. 2008;49(12):1304–12.
42. Ungdata BA. Nasjonale resultater 2018. NOVA: Oslo; 2018.
43. Abebe DS, Lien L, Hjelde KH. What we know and Don't know about mental health problems among immigrants in Norway. J Immigr Minor Health. 2014;16(1):60–7.
44. Bøe T, Øverland S, Lundervold AJ, Hysing M. Socioeconomic status and children's mental health: results from the Bergen child study. Soc Psychiatry Psychiatr Epidemiol. 2012;47(10):1557–66.
45. Erola J, Jalonen S, Lehti H. Parental education, class and income over early life course and children's achievement. Res Soc Stratification Mobility. 2016; 44:33–43.
46. Youthinmind. Information for researchers and professionals about the Strengths & Difficulties Questionnaires 2018. Available from: http://www.sdqinfo.com/.
47. Altman DG, Bland JM. Missing data. BMJ (Clinical research ed). 2007; 334(7590):424.
48. Rubin DB. Multiple imputation after 18+ years. J Am Stat Assoc. 1996; 91(434):473–89.
49. Pedersen AB, Mikkelsen EM, Cronin-Fenton D, Kristensen NR, Pham TM, Pedersen L, et al. Missing data and multiple imputation in clinical epidemiological research. Clin Epidemiol. 2017;9:157–66.
50. Manly CA, Wells RS. Reporting the use of multiple imputation for missing data in higher education research. Res High Educ. 2015;56(4):397–409.
51. Sterne JAC, White IR, Carlin JB, Spratt M, Royston P, Kenward MG, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. BMJ. 2009;338:b2393.
52. Molner-Urdiales D, Ruiz JR, Vicente-Rodriguez G, Ortega FB, Rey-Lopez JP, España-Romero V, et al. Associations of muscular and cardiorespiratory fitness with total and central body fat in adolescents: The HELENA Study. Br J Sports Med. 2011;45(2):101–8.
53. Andersen JR, Natvig GK, Aadland E, Moe VF, Kolotkin RL, Andersen SA, et al. Associations between health-related quality of life, cardiorespiratory fitness, muscle strength, physical activity and waist circumference in 10-year-old children: the ASK study. Qual Life Res. 2017;26(12):3421–8.
54. Rieck T, Jackson A, Martin SB, Petrie T, Greenleaf C. Health-related fitness, body mass index, and risk of depression among adolescents. Med Sci Sports Exerc. 2013;45(6):1083–8.
55. Kelly NR, Mazzeo SE, Evans RK, Stern M, Thacker LF, Thornton LM, et al. Physical activity, fitness and psychosocial functioning of obese adolescents. Ment Health Phys Act. 2010;4(1):31–7.
56. Padilla-Moledo C, Ruiz JR, Ortega FB, Mora J, Castro-Pinero J. Associations of muscular fitness with psychological positive health, health complaints, and health risk behaviors in Spanish children and adolescents. J Strength Cond Res. 2012;26(1):167–73.
57. Kettunen O, Kyröläinen H, Santtila M, Vuorimaa T, Vasankari TJ. Greater levels of cardiorespiratory and muscular fitness are associated with low stress and high mental resources in normal but not overweight men. BMC Public Health. 2016;16:788.
58. Altman DG, Royston P. The cost of dichotomising continuous variables. BMJ (Clinical research ed). 2006;332(7549):1080.
59. Fiatarone MA, Marks EC, Ryan ND, Meredith CN, Lipsitz LA, Evans WJ. High-intensity strength training in nonagenarians: effects on skeletal muscle. JAMA. 1990;263(22):3029–34.
60. Choi KW, Chen C-Y, Stein MB, Klimentidis YC, Wang M-J, Koenen KC, et al. Assessment of bidirectional relationships between physical activity and depression among adults: a 2-sample mendelian randomization study. JAMA Psychiatry. 2019;76(4):399–408.
61. Parfitt G, Pavey T, Rowlands AV. Children's physical activity and psychological health: the relevance of intensity. Acta Paediatr. 2009;98(6):1037–43.
62. Ahn S, Fedewa AL. A meta-analysis of the relationship between Children's physical activity and mental health. J Pediatr Psychol. 2011;36:385–97.

14. Blair SN, Cheng Y, Holder JS. Is physical activity or physical fitness more important in defining health benefits? Med Sci Sports Exerc. 2001;33(6 Suppl):379–99.
15. Caspersen CJ, Powell KE, Christenson GM. Physical activity, exercise and physical fitness: definitions and distinctions for health-related research. Public Health Rep. 1985;100(2):126–30.
16. Silverman MN, Deuster PA. Biological mechanisms underlying the role of physical fitness in health and resilience. Interface Focus. 2014;4(5):20140040.
17. Voss MW, Weng TB, Burzynska AZ, Wong CN, Cooke GE, Clark R, et al. Fitness, but not physical activity, is related to functional integrity of brain networks associated with aging. Neuroimage. 2016;131:113–25.
18. Silva G, Andersen LB, Aires L, Mota J, Oliveira J, Ribeiro JC. Associations between sports participation, levels of moderate to vigorous physical activity and cardiorespiratory fitness in children and adolescents. J Sports Sci. 2013;31(12):1359–67.
19. Streeter VM, Milhausen RR, Buchholz AC. Body image, body mass index, and body composition in young adults. Can J Diet Pract Res. 2012;73(2):78–83.
20. Grabe S, Ward LM, Hyde JS. The role of the media in body image concerns among women: a meta-analysis of experimental and correlational studies. Psychol Bull. 2008;134(3):460–76.
21. Luppino FS, de Wit LM, Bouvy PF, Stijnen T, Cuijpers P, Penninx BWJH, et al. Overweight, obesity, and depression: a systematic review and meta-analysis of longitudinal studies. Arch Gen Psychiatry. 2010;67(3):220–9.
22. Rodgers RF, Ganchou C, Franko DL, Chabrol H. Drive for muscularity and disordered eating among French adolescent boys: a sociocultural model. Body Image. 2012;9(3):318–23.
23. O'Connor PJ, Herring MP, Caravalho A. Mental health benefits of strength training in adults. Am J Lifestyle Med. 2010;4(5):377–96.
24. Schuch FB, Vancampfort D, Sui X, Rosenbaum S, Firth J, Richards J, Ward PB, Stubbs B. Are lower levels of cardiorespiratory fitness associated with incident depression? A systematic Review of Prospective Cohort Studies. Prev Med. 2016;93:159–65.
25. Vasiliki K, Mamadjanov T, Meisinger C, Linseisen J. Association between muscular strength and depressive symptoms. Wien Klin Wochenschr. 2019;131(11):255–64.
26. Xu Q, Anderson D, Lurie-Beck J. The relationship between abdominal obesity and depression in the general population: A systematic review and meta-analysis. Obes Res Clin Pract. 2011;5(4):e267–78.
27. Ruggero CJ, Petrie T, Sheinbein S, Greenleaf C, Martin S. Cardiorespiratory fitness may help in protecting against depression among middle school adolescents. J Adolesc Health. 2015;57(1):60–5.
28. Greenleaf C, Petrie T, Martin SB. Psychosocial variables associated with body composition and cardiorespiratory fitness in middle school students. Res Q Exerc Sport. 2010;81:65–74.
29. Lubans DR, Cliff DP. Muscular fitness, body composition and physical self-perception in adolescents. J Sci Med Sport. 2011;14(3):216–21.
30. Smith JJ, Eather N, Morgan PJ, Plotnikoff RC, Faigenbaum AD, Lubans DR. The health benefits of muscular fitness for children and adolescents: a systematic review and meta-analysis. J Sports Medicine. 2014;44(9):1209–23.
31. Yeatts PE, Martin SB, Petrie TA. Physical fitness as a moderator of neuroticism and depression in adolescent boys and girls. Personal Individ Differ. 2017;114:30–5.
32. Haugen T, Ommundsen Y, Seiler S. The relationship between physical activity and physical self-esteem in adolescents: the role of physical fitness indices. Pediatr Exerc Sci. 2013;25(1):138–53.
33. Kolle E, Solberg RB, Grydeland M, Sälvenporn R, Andersen SA, Ekelund U, Berntsen S, Resaland GK, Dyrstad SM, Steene-Johannessen J. Evaluating the effect of increased physical activity in lower secondary school adolescents: The study protocol for the School in Motion (SciM) study – a cluster-randomized controlled trial. Manuscript submitted for publication; 2020.
34. Council of Europe. Handbook of the Eurofit tests of Physical Fitness. Roma Italian National. 1988.
35. Andersen L, Andersen TE, Andersen E, Andersen S. An intermittent running test to estimate maximal oxygen uptake: the Andersen test. J Sports Med Phys Fitness. 2008;48(4):434–7.
36. Goodman R. The strengths and difficulties questionnaire: a research note. J Child Psychol Psychiatry. 1997;38(5):581–6.
37. Goodman A, Goodman R. Strengths and difficulties questionnaire as a dimensional measure of child mental health. J Am Acad Child Adolesc Psychiatry. 2009;48(4):400–3.
38. Rothenberger A, Becker A, Erhart M, Wille N, Ravens-Sieberer U. Psychometric properties of the parent strengths and difficulties questionnaire in the general population of German children and adolescents: results of the BELLA study. Eur Child Adolesc Psychiatry. 2008; 17(1):99.
39. Goodman R. Psychometric properties of the strengths and difficulties questionnaire. J Am Acad Child Adolesc Psychiatry. 2001;40(11):1337–45.
40. Muris P, Meesters C, Eijkelenboom A, Vincken M. The self-report version of the strengths and difficulties questionnaire: its psychometric properties in 8- to 13-year-old non-clinical children. Br J Clin Psychol. 2004;43:437–48.
41. Van Roy B, Veenstra M, Clench-Aas J. Construct validity of the five-factor strengths and difficulties questionnaire (SDQ) in pre-, early, and late adolescence. J Child Psychol psychiatry. 2008;49(12):1304–12.
42. Ungdata BA. Nasjonale resultater 2018. NOVA: Oslo; 2018.
43. Abebe DS, Lien L, Hjelde KH. What we know and Don't know about mental health problems among immigrants in Norway. J Immigr Minor Health. 2014;16(1):60–7.
44. Bøe T, Øverland S, Lundervold AJ, Hysing M. Socioeconomic status and children's mental health: results from the Bergen child study. Soc Psychiatry Psychiatr Epidemiol. 2012;47(10):1557–66.
45. Erola J, Jalonen S, Lehti H. Parental education, class and income over early life course and children's achievement. Res Soc Stratification Mobility. 2016; 44:33–43.
46. Youthinmind. Information for researchers and professionals about the Strengths & Difficulties Questionnaires 2018. Available from: http://www.sdqinfo.com/.
47. Altman DG, Bland JM. Missing data. BMJ (Clinical research ed). 2007; 334(7590):424.
48. Rubin DB. Multiple imputation after 18+ years. J Am Stat Assoc. 1996; 91(434):473–89.
49. Pedersen AB, Mikkelsen EM, Cronin-Fenton D, Kristensen NR, Pham TM, Pedersen L, et al. Missing data and multiple imputation in clinical epidemiological research. Clin Epidemiol. 2017;9:157–66.
50. Manly CA, Wells RS. Reporting the use of multiple imputation for missing data in higher education research. Res High Educ. 2015;56(4):397–409.
51. Sterne JAC, White IR, Carlin JB, Spratt M, Royston P, Kenward MG, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. BMJ. 2009;338:b2393.
52. Moliner-Urdiales D, Ruiz JR, Vicente-Rodriguez G, Ortega FB, Rey-Lopez JP, España-Romero V, et al. Associations of muscular and cardiorespiratory fitness with total and central body fat in adolescents: The HELENA Study. Br J Sports Med. 2011;45(2):101–8.
53. Andersen JR, Natvig GK, Aadland E, Moe VF, Kolotkin RL, Andersen SA, et al. Associations between health-related quality of life, cardiorespiratory fitness, muscle strength, physical activity and waist circumference in 10-year-old children: the ASK study. Qual Life Res. 2017;26(12):3421–8.
54. Rieck T, Jackson A, Martin SB, Petrie T, Greenleaf C. Health-related fitness, body mass index, and risk of depression among adolescents. Med Sci Sports Exerc. 2013;45(6):1083–8.
55. Kelly NR, Mazzeo SE, Evans RK, Stern M, Thacker LF, Thornton LM, et al. Physical activity, fitness and psychosocial functioning of obese adolescents. Ment Health Phys Act. 2010;4(1):31–7.
56. Padilla-Moledo C, Ruiz JR, Ortega FB, Mora J, Castro-Pinero J. Associations of muscular fitness with psychological positive health, health complaints, and health risk behaviors in Spanish children and adolescents. J Strength Cond Res. 2012;26(1):167–73.
57. Kettunen O, Kyröläinen H, Santtila M, Vuorimaa T, Vasankari TJ. Greater levels of cardiorespiratory and muscular fitness are associated with low stress and high mental resources in normal but not overweight men. BMC Public Health. 2016;16:788.
58. Altman DG, Royston P. The cost of dichotomising continuous variables. BMJ (Clinical research ed). 2006;332(7549):1080.
59. Fiatarone MA, Marks EC, Ryan ND, Meredith CN, Lipsitz LA, Evans WJ. High-intensity strength training in nonagenarians: effects on skeletal muscle. JAMA. 1990;263(22):3029–34.
60. Choi KW, Chen C-Y, Stein MB, Klimentidis YC, Wang M-J, Koenen KC, et al. Assessment of bidirectional relationships between physical activity and depression among adults: a 2-sample mendelian randomization study. JAMA Psychiatry. 2019;76(4):399–408.
61. Parfitt G, Pavey T, Rowlands AV. Children's physical activity and psychological health: the relevance of intensity. Acta Paediatr. 2009;98(6):1037–43.
62. Ahn S, Fedewa AL. A meta-analysis of the relationship between Children's physical activity and mental health. J Pediatr Psychol. 2011;36:385–97.

Paper 2

The current issue and full text archive of this journal is available on Emerald Insight at:
https://www.emerald.com/insight/0965-4283.htm

The process evaluation of a school-based physical activity intervention: influencing factors and potential consequences of implementation

Received 12 January 2020
Revised 1 April 2020
Accepted 21 April 2020

Andreas Åvitsland, Stein Erik Ohna and Sindre Mikal Dyrstad
University of Stavanger, Stavanger, Norway
Hege Eikeland Tjomsland and Øystein Lerum
Western Norway University of Applied Sciences, Bergen, Norway, and
Eva Leibinger
University of Stavanger, Stavanger, Norway

Abstract

Purpose – This paper evaluates the implementation of a school-based physical activity intervention and discusses how the intervention outcomes can be influenced by the implementation.

Design/methodology/approach – In four of the nine lower secondary schools in which the intervention was conducted, the authors examined implementation fidelity, adaptation, quality, responsiveness and dose received. The authors conducted focus group interviews with teachers ($n = 8$) and students ($n = 46$) and made observations. Dose delivered was examined quantitatively, with weekly registrations.

Findings – Results showed that two out of four schools made few and positive adaptations, implemented the intervention with high fidelity and quality and responded positively. Four main factors were found to influence implementation: frame factors, intervention characteristics, participant characteristics and provider characteristics.

Research limitations/implications – A cross-sectional design was used and may not represent implementation throughout the whole school year.

Practical implications – In terms of large-scale implementation, the intervention may be generalizable. However, intervention criteria such as adequate facilities and a flexible timetable may be unattainable for some schools. The intervention can be adapted without compromising its purpose, but adaptations should be a result of cooperation between students and teachers.

Originality/value – Process evaluations on this topic are rare. This study adds to a limited knowledge base concerning what factors may influence implementation of school-based physical activity interventions for adolescents.

Keywords Process evaluation, School, Intervention, Physical activity, Implementation, Qualitative methods
Paper type Research paper

Introduction

Schools are considered viable settings for intervention, reaching children and adolescents irrespective of sex, socioeconomic status and ethnicity. Studies evaluating school-based physical activity (PA) interventions have largely examined whether the interventions affect

Health Education
Emerald Publishing Limited
0965-4283
DOI 10.1108/HE-01-2020-0004

students' academic achievement (Norris *et al.*, 2015), physical activity and fitness (Kriemler *et al.*, 2011) or mental health (Smith *et al.*, 2018; Lubans *et al.*, 2016). The way in which such interventions are implemented may partly determine their effectiveness (Durlak and Dupre, 2008). However, monitoring and evaluating implementations has not been prioritized in school-based PA intervention research (Naylor *et al.*, 2015; Watson *et al.*, 2017; Daly-Smith *et al.*, 2018), increasing the risk of a type 3 error (Dobson and Cook, 1980). A type 3 error can occur when researchers dismiss a potentially effective intervention based on unsatisfactory results, when, in fact, the results were caused by poor implementation and not the intervention itself. Among the existing studies evaluating school-based PA interventions, only a few have addressed the link between implementation and outcome (Naylor *et al.*, 2015), thus limiting our understanding of their relationship.

Recent reviews of school-based PA interventions have shown mixed results in terms of physical, mental and cognitive outcomes (Demetriou and Honer, 2012; Daly-Smith *et al.*, 2018; Watson *et al.*, 2017; Singh *et al.*, 2018; Hynynen *et al.*, 2016). One reason for this may have to do with the difference between efficacy and effectiveness. Efficacy refers to an intervention's effect under ideal circumstances, while effectiveness refers to an intervention's effect under real-life conditions (Revicki and Frank, 1999). Conducting interventions in real-life conditions, such as in schools, makes it more difficult to assess effects, since variables outside the researcher's control may come into play. Additionally, schools are complex social systems (Moore *et al.*, 2019), because the individual agents (e.g. students, teachers) within the system are "numerous, dynamic, autonomous, highly interactive, learning and adaptive" (Keshavarz *et al.*, 2010). These agents and other school factors can influence implementation (Clarke, 2010). For instance, in their review of school-based PA interventions, Naylor *et al.* (2015) identified 20 factors that could hamper implementation. "Limited time" was the most frequently mentioned factor, followed by "lack of resources," "lesson scheduling" and "weather." The outcomes from a school-based cluster randomized controlled trial (RCT) may, therefore, have more to do with the specific school context than the intervention itself.

Interventions can also be complex, posing further challenges to evaluating outcomes (Craig *et al.*, 2008). When a complex intervention is introduced in several unique complex contexts, dynamic processes and interactions can lead to the intervention evolving differently in the different contexts. This could in turn influence the intervention outcomes (Hawe *et al.*, 2009) and obstruct researchers' ability to make causal inferences (Rickles, 2009). However, by monitoring and evaluating the process of implementation, we can improve our ability to interpret outcomes and causes (Oakley *et al.*, 2006; Tarp *et al.*, 2016).

This paper reports on the process evaluation of the "Don't worry, be happy" (DWBH) intervention, one of two separate intervention arms in a cluster RCT called School in Motion (ScIM). ScIM was designed to assess whether 120 min of additional weekly school time PA affected students' academic achievement, learning environment, physical fitness, PA levels and mental health. The findings of the process evaluation should be used as a supplementary tool when interpreting the outcomes of the ScIM study, which will be published in upcoming papers.

Aim and research questions

The aim of this research was to evaluate the implementation of the DWBH intervention. Furthermore, in the context of searching for feasible methods to increase school time PA for adolescents, we aim to increase our understanding of what works under what circumstances and why. In order to reach these aims, we asked the following questions:

(1) How was the intervention implemented?

(2) What influenced the implementation?

Theoretical background

The purpose of a process evaluation is to "explore the implementation, receipt, and setting of an intervention and help in the interpretation of the outcome results" (Oakley *et al.*, 2006). The DWBH intervention can be seen as a complex intervention, which according to Craig *et al.* (2008), is characterized by (1) the number of interacting components, (2) the number and difficulty of behaviors required by those delivering or receiving the intervention, (3) the number of groups or organizational levels targeted by the intervention, (4) the number and variability of outcomes and (5) the degree of flexibility. The present study understands implementation as defined by Durlak (2016): ". . .the ways a programme is put into practice and delivered to participants." According to Durlak and Dupre (2008), implementation consists of eight separate but overlapping aspects. The present study addresses five: fidelity, adaptation, quality, responsiveness and dose delivered. Dose received does not fall under implementation as defined by Durlak (2016) but has been included as a sixth aspect in this evaluation to help understand how relevant the intervention is to the target participants (Linnan and Steckler, 2002). While dose delivered refers to the specific amount of intervention that has been provided to its participants, dose received refers to actual participation or attendance (Berkel *et al.*, 2011). Including both can offer important information about their relationship.

Besides dose received, focusing on the aforementioned aspects draws on the argumentation by Berkel *et al.* (2011), that they "occur within the delivery of program sessions, and as a result, constitute potential sources of disconnect between the program as designed and that which is implemented". While fidelity, adaptation, quality and dose delivered are mainly determined by the intervention providers, responsiveness and dose received are determined by the participants, allowing us to examine the dynamic relationship between provider and participant, who can influence each other and ultimately, the intervention results (Berkel *et al.*, 2011). Evaluating quality has been neglected in previous implementation evaluations of school-based PA interventions (Naylor *et al.*, 2015). Humphrey *et al.* (2017) argue, however, that quality may be more important than fidelity and dose when it comes to impacting study outcomes; so this aspect was included. The other three aspects of implementation – differentiation, monitoring control and reach – can also influence the outcomes, but are not explicitly part of the program delivery or the relationship between provider and participant. We therefore chose not to focus on these aspects.

Methods

The School in Motion study

The ScIM study was a multicenter trial conducted in Norway during the 2017–2018 school year. The study was initiated by the Norwegian Directorate of Health and the Directorate of Education and Training. The interventions were designed by the project management group at the Norwegian School of Sports Sciences, piloted with a small number of classes and adjusted, before commencing the ScIM study. The project management group invited 103 lower secondary schools to participate, and 29 schools accepted. Only students attending ninth grade (14–15 years) during the intervention period were included. The intervention period lasted 29 weeks and since the intervention would be part of the school's curriculum, all the 2,733 eligible students were required to participate, as they would in any other school subject. The DWBH intervention was assigned to nine schools (663 eligible students). Each school chose one teacher liaison, who was responsible for the intervention execution and for communicating with the researchers at their respective test center. The ScIM study was approved by the Norwegian Centre for Research Data (project number 49094).

The Don't worry, be happy intervention. The goal of the DWBH intervention is to facilitate positive experiences with PA in order to contribute to healthy adolescent development.

A summary of key DWBH intervention characteristics and their relationship with fidelity, adaptation, quality, responsiveness and dose delivered/received are presented in Table 1. DWBHs theoretical framework and components are extensively described elsewhere (Kolle *et al.*, 2020). Briefly, DWBH draws on the theoretical perspectives on positive youth development (Lerner, 2015), relational developmental systems theory (Lerner *et al.*, 2015) and positive movement experiences (Agans *et al.*, 2013). According to the intervention's underlying theories, the goal of the DWBH intervention is achieved by allowing the students individuality and personal interest in their choice of activities. Furthermore, the activities need to take place in a social environment that enables the participants to experience competence, confidence, connection, character and caring.

The intervention consisted of two separate weekly lessons, "Don't worry" (DW) and "Be happy" (BH). DW could be conducted as a regular physical education (PE) lesson, although we encouraged the teachers to allow students to pursue activities of their choice. BH was the main component and requires a more detailed description. There was an initial planning phase in which students were asked to choose an activity or sport they wanted to pursue for the rest of the school year. Next, students who chose the same activity formed groups and started planning long-term activity goals, conflict solution strategies and organized a leadership structure. The plans were formalized in a written "activity contract" (see attachment 1), which was signed by the group members and approved by the teacher. Once approved, the planning phase was over, and the students pursued their activities in the BH lessons for the rest of the school year. Students were not assessed in BH. Although BH was student-led, teachers were the formal providers of the intervention and were required to be qualified PE teachers. Their tasks were to be present, observe and provide guidance when necessary. BH was to be organized so that students could participate together across homeroom classes.

Guidelines were provided for the incorporation of the two DWBH lessons into the school's schedule. To schedule one of the lessons, the schools were required to reallocate 5% of the time from other subjects. The other lesson was added on top of the existing schedule. This

Table 1. Summary of key intervention characteristics and their relationship with the implementation aspects

Fidelity	(1)	"Don't worry": Similar to ordinary PE in separate classes but students have the opportunity to choose their activity. "Be happy": students pursue activities of their choice, in groups across homeroom classes, that they formed themselves
	(2)	Students choose activity based on interest, not based on who they can be together within the group. Groups should stay together. Maximum eight students per group
	(3)	Sufficient information must be given to the students. Students must use the activity contract to conduct long-term planning. All groups must have a leadership structure, group goals and plan for conflict resolution
	(4)	5% of time from other subjects should be taken to make room for one lesson. The second comes in addition to the ordinary schedule
Quality	(1)	Sufficient facilities and equipment
	(2)	Sufficient teacher-to-student ratio
	(3)	Teachers should be available and able to help when necessary. They should interfere when necessary by recognizing the need for flexibility, evaluation, group alterations and conflict resolution
Responsiveness	(1)	Positive response toward DWBH, and regarding the intervention as relevant, useful, advantageous in any way
	(2)	Responsiveness also applies to how the teachers respond to the intervention
Dose delivered/ received	(1)	Two lessons per week during the 29-week intervention period
Adaptations	(1)	No predetermined adaptations have been defined. Small adaptations can be made if necessary and/or if they benefit overall implementation

resulted in students having a 45–60 min longer school day once a week, for which the schools were economically compensated. The schools were free to choose when to conduct the two lessons but were encouraged to choose two separate days of the week.

The process evaluation

Design and participants. This process evaluation used a cross-sectional design to gather qualitative data by conducting interviews and observations. A longitudinal design was used to collect quantitative data, by teachers reporting dose delivered each week throughout the intervention period. Participant sampling for the interviews was a combination of random and purposive. We randomly selected four schools assigned DWBH, one from each of the four regions in Norway where the intervention was carried out. School 1 included 52 students from two classes and was located in a rural area outside a major city. School 2 included 117 students from four classes and was located in a residential/rural area outside a major city. School 3 included 47 students from two classes and was located in a residential/urban area close to a smaller city. School 4 included 87 students from four classes and was located in a residential/urban area between two moderate-sized cities. Teacher liaisons at each school accepted the invitation to participate in the process evaluation. Next, the teacher liaisons were asked to perform a purposive sampling of students to be interviewed: three activity groups representing different activities and opinions toward the intervention and PA in general. The purposive sampling strategy was employed to cover a diversity of experience, in order to prevent bias toward presenting only one type of information from one type of participants (Robinson, 2014). Teachers who supervised DWBH were also interviewed to obtain knowledge about the implementation process from the providers' perspectives. A total of 54 individuals were interviewed. This amounted to 12 student focus group interviews ($n = 46$), two individual teacher interviews and two teacher focus group interviews ($n = 6$). Student interviewees provided written informed consent from their parents, and teacher interviewees gave their consent verbally. This study has been designed, conducted and reported so as to ensure the confidentiality and anonymity of participants.

Data collection. We conducted semistructured focus group interviews and individual interviews to capture participants' and providers' experiences of the intervention. Initially, all the interviews were supposed to be conducted in groups, but two teacher interviews were individual because of illness among the teachers. The interview guide was constructed to elicit answers that could be linked to the six included implementation aspects (see attachment 2). We chose semistructured interviews based on our expectation of broad variation of opinions and experiences regarding DWBH and other emerging issues (Barriball and While, 1994). Focus group interviews are suitable for program evaluation because participants answer questions about a specific topic in a social context; they can discuss and potentially reveal information that would not have emerged in an individual interview (Frey and Fontana, 1991).

Secondary data were gathered (Manzano, 2016) by observing one BH lesson in each of the selected schools. Observations were conducted to experience a physical presence, which provides an impression of the environmental surroundings, the participants, attitude toward the intervention and the dynamics between provider and participants. The purpose of these impressions was to assist in analyzing the interviews.

Qualitative data were gathered within the same week in each school, during the second half of the intervention period (between mid-January and mid-March 2018). The interviews took place during school hours at the participants' respective schools, in a classroom with only the researcher and the interviewee(s) present. The interviews lasted between 30 and 55 min and were audio recorded.

We defined dose delivered as the percentage of lessons provided to the participants, relative to the number of lessons that were possible to provide during the intervention period.

To measure dose delivered, the teacher liaisons used an online registration tool to weekly register DWBH lessons as executed/not executed. We considered 80% and above to be a high delivered dose. It is important to note that dose received is also a quantitative concept and by assessing it qualitatively, we only get an indication of how attendance has been relative to the dose delivered.

Data analysis. Audio recordings of the interviews were imported into NVivo qualitative data analysis Software 12 (QSR International Pty Ltd., Doncaster, Australia) and transcribed verbatim. Data were further analyzed in NVivo and Excel, using the five steps of the framework analysis (Spencer and Ritchie, 2002). The rationale behind choosing this approach is that it exists within the family of content or thematic analysis and can be used to "identify commonalities and differences (. . .) focusing on relationships between different parts of the data, thereby seeking to draw descriptive and/or explanatory conclusions" (Gale *et al.*, 2013). How we followed the five steps is outlined in Figure 1. Briefly, one deductive analysis and one inductive analysis were conducted consecutively. The deductive analysis was guided by research question 1 (how the intervention was implemented), while the inductive analysis was guided by research question 2 (what influenced implementation). The first author (AA) was responsible for coding the material. The codes and initial analyses were discussed in meetings with three of the coauthors (SEO, SD, EL), and this contributed to interpreting, summarizing and synthesizing the data. The inductive analysis resulted in the merging of

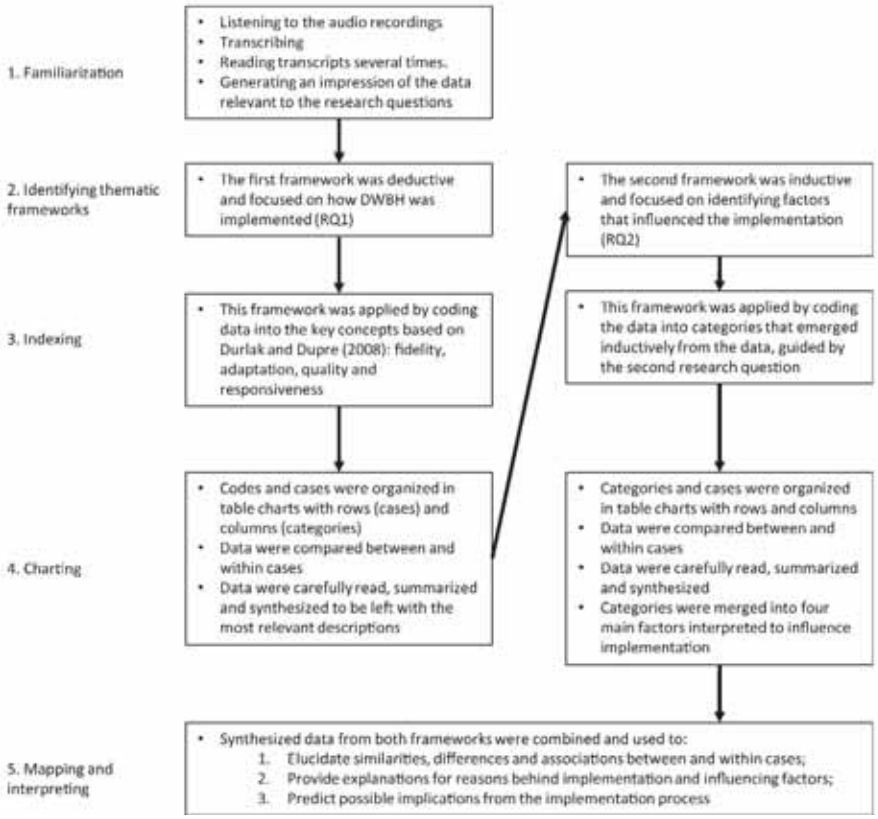


Figure 1.
Detailed outline of how the five steps of the framework analysis were followed

subcategories into four main factors that were interpreted to be the influencers of implementation: (1) frame factors, (2) intervention characteristics, (3) participant characteristics and (4) provider characteristics (see Table 2). The fifth and final step of the analysis (mapping and interpreting) involved combining the inductive and deductive findings to interpret and outline processes behind the implementation and the influencing factors. Notes from the observations assisted in the interpretation and coding of the interview material, in particular by supporting the coherence between what the interviewees stated and what was observed (Mays and Pope, 1995).

Results

In the following section, results are presented within the context of the main factors that were found to influence fidelity, adaptation, quality, responsiveness and dose delivered/received. Average dose delivered in the five schools that were not included in the process evaluation was 80%.

Frame factors

Frame factors represent the contextual opportunities and limitations that exist on an organizational and environmental level. As an influencer to all the examined implementation aspects, scheduling was one of the most influential frame factors. Schools 1 and 4 scheduled DWBH as intended: one lesson was added to the existing schedule, while the other lesson

Main factors	Subcategories	Sample quotes
Frame factors	(1) Adapting intervention components (2) After school consequences of DWBH (3) Facilities and equipment (4) Scheduling (5) Teacher availability	Subcategory 3, 4 and 5 *We do not have access to the gymnasium that one lesson [DW] and we do not have a teacher either, just a substitute teacher sometimes, so we just end up playing cards.* (School 3, students)
Intervention characteristics	(1) Freedom of choice (2) All classes together (3) No assessment (4) Extra time in school	Subcategory 1 *I think they [students] think that this [DWBH] is very enjoyable, particularly that they are allowed to decide themselves what to do.* (School 4, teachers)
Participant characteristics	(1) Characteristics, skills, other (2) Efforts, engagement, interest and motivation (3) Group dynamics, relationships and conflicts	Subcategory 1 and 2 *How they [students] handle the freedom in DWBH is about maturity. Those who pursue activities in school do what they are supposed to, but some, who probably could be in school, choose something else, off school grounds because it is easier to sneak off.* (School 2, teachers)
Provider characteristics	(1) Actions, involvement, supervision, guidance and communication (2) Characteristics, attitudes and motivation (3) Belief in students' capabilities (4) Participating in student activities	Subcategory 1 *That time when our plans just got out of control, the teacher came and gave us some directions, so we changed the group and changed our plan; he kind of straightened us out.* (School 1, students)

Table 2.
Factors influencing the implementation of "Don't worry, be happy," with merged subcategories and sample quotes

received 5% allocated time from other subjects. School 2 was unable to schedule DWBH as intended and added both lessons on top of the existing schedule, giving students two additional periods in school per week. This had undesirable consequences for the students because they were unable to take the school bus after the extra lessons. Many of these students had long travel distances to school and missing the school bus meant that many arrived at home much later than usual. These consequences strongly influenced how students in School 2 responded to the intervention, illustrated by the following quotes:

> School 2, group 1:
>
> Student 1: We do not get a school bus, that's the worst part, I think.
>
> S2: We could have gotten the bus home, because it gets really stressful.
>
> S1: Yes, that's the least they could have done.
>
> (. . .)
>
> S2: We spend a long time to get home (. . .) it makes us late for sports practice.
>
> S3: It messes up everything.

School 3 was also unable to schedule the intervention as intended: one subject was removed, and time was allocated from one other subject and one recess period. Although the adaptation limited recess time, the students were still positive toward the intervention.

Schools 1, 2 and 3 scheduled the two lessons on separate days, as the intervention guidelines recommended, while School 4 scheduled DWBH as one double period. The students preferred this adaptation, and according to the teachers, it was necessary:

> School 4, teachers:
>
> Teacher 1: Because the gymnasium is a ways away, we would have lost a lot of time if we had two single lessons.
>
> T2: The lessons would have had to be the last of the day (. . .) so they had time to change clothes before and after. That takes a lot of time.

Facilities emerged as an important factor influencing fidelity, quality, responsiveness and dose received. There were big differences between schools: the gymnasium in School 3 was too small to accommodate two classes at once; therefore, the teacher made the adaptation to have both lessons with separate classes. This scheduling, however, in combination with limited facilities and limited teacher availability, caused another problem for one of the two classes: During DW, they did not have access to the gymnasium because it was being used by another class, and they did not have a teacher because one of the two supervising teachers was on long-term sick leave. Sometimes, a substitute teacher would be present, but they were rarely aware of the purpose of the lesson. The dose delivered in School 3 was registered to be 75%, but these limiting factors suggest that the dose received might be lower:

> School 3, teacher:
>
> It was a bit embarrassing last time, because the local ScIM-coordinator came from the university to observe DW, and the substitute had no clue about what was supposed to happen. So the students had just said "well, usually we just play cards in these lessons", so they were just inside the classroom playing cards, which wasn't good.

Schools 1 and 4 had spacious facilities that allowed all the students to participate together. These students expressed satisfaction with the facilities, which included large gymnasiums with ample equipment and many opportunities for outdoor activities. The teachers from

School 2, however, said that their facilities were too limited to carry out DWBH as intended. The gymnasium was small and there was a swimming pool that, according to the students, was often closed for maintenance. The limited facilities led to the adaptation that two-thirds of students had to pursue their activities off school grounds. Additionally, the number of teachers available was limited, as two teachers were always assigned to the swimming pool for safety reasons. These combined factors made it difficult to see and supervise the students on and off school grounds. In contrast, there were always four teachers present during DWBH in School 4. This made it possible to supervise all students inside and outside the gymnasium. The teachers sometimes felt superfluous, but were cognizant of the advantages of being many as illustrated by the following excerpt:

> School 4, teachers:
>
> T3: . . .after all, they're 80 students
>
> T2: Yes, and they're spread out, outside and inside
>
> T3: They're everywhere, so if the fire alarm goes off, two teachers will not be enough (. . .) so we might feel superfluous then and there, but it is a safety factor.
>
> T2: Mhm, and when we're four, it's easier to supervise groups well, we can be outside by the soccer field, we can go to the gym down the street. . .
>
> T3: . . .where we have students, and there's a hiking group we can join, so there are more options

Adaptations to the intervention emerged as a factor influencing fidelity, quality and responsiveness. For instance, responsiveness was positively influenced by the adaptation to execute DW in the same way as the BH lesson, which occurred in Schools 1, 3 and 4. In School 2, however, DW was usually organized as what resembled a PE lesson, in which the students could vote on an activity. The teacher justified this adaptation by contending that he had difficult students who did not take DW seriously and needed stricter boundaries. In one class, the vote was often soccer, which left some students unhappy with the lesson:

> School 2, group 2:
>
> S1: We have a vote, but the problem is that it always ends up with soccer (. . .) I'm not so fond of soccer; it's okay, but some people hate soccer, so not so many like DW.

Intervention characteristics

Intervention characteristics represent specific components of the intervention, such as additional time in school, additional PA, freedom to choose and the lack of assessment. Various intervention characteristics emerged as factors that influenced responsiveness among students and teachers. Students in Schools 1, 3 and 4 most frequently mentioned the freedom to pursue an activity that interested them as a positive intervention characteristic:

> School 1, group 1:
>
> S1: We are interested in it, so you will not get that "oh I do not want to do this" or whatever.
>
> S3: It's like, we do it properly because we like it and then we want to do it.

The students in School 2 did not care that they were able to choose their own activity, because they were not interested in having the intervention at all:

> School 2, group 1:
>
> S2: To me, it was just random. I just chose something, to have something to do, really.

(. . .)

S1: We did not really want to do it (laughter)

Additional time in school also potentially influenced responsiveness. As previously mentioned, students in School 2 received twice as much additional time as DWBH intended, which caused negative responses. Conversely, the students in Schools 1, 3 and 4 received just one additional period. Some of these students took issue with the added time, but also said it was worth it because they could have DWBH.

An important intervention characteristic was that students were not supposed to be assessed during BH. Teachers in Schools 1 and 3, and two student groups, respectively from Schools 3 and 4 talked positively about the lack of assessment. In contrast, a student group from School 2 would have preferred to have the lessons assessed:

School 2, group 2:

S2: It's better if we can, like. . .

S1: Use it for something

S2: Show it, yes, use it for something, like, if you do a great job in BH, none of the teachers see it (. . .) it's better if it matters for the grade, and it's not just something we do, like, for nothing, since everything is about grades in lower secondary school (laughter).

The students' freedom of choice and the absence of assessment were intervention characteristics that influenced teachers' responsiveness positively in Schools 1, 3 and 4. In contrast, teachers in School 2 acknowledged that DWBH had some good ideas, but argued against some of the intervention's characteristics, such as having BH for all classes simultaneously and allowing students excessive freedom to choose. The following quote illustrates a somewhat negative attitude toward the intervention and the researchers who designed it:

School 2, teachers:

T1: I think the whole thing shows that DWBH was designed by people who do not work in school, because there are a lot of good intentions but when you face common practice, it becomes difficult.

Participant characteristics

Participant characteristics include participants' attitudes, skills, interests, actions and participation that are specifically related to the intervention. Students' attitude toward the intervention was likely an important influencing factor for dose received: all activity groups in Schools 1, 3 and 4 repeatedly stated that they enjoyed DWBH and all groups wanted to continue with the lessons in tenth grade. The delivered doses reported from Schools 1 and 4 were 81 and 86%, respectively, and the positive attitudes expressed in the interviews indicate that the received dose was high, that is, truancy was low. School 2 registered 75% of the dose delivered; however, the interviews indicate that the received dose might have been lower, because of truancy. The negative attitudes toward DWBH suggested high motivation for truancy. Consequently, when BH was scheduled as the last lesson of the day and when two-thirds of students could leave school to do their activities elsewhere, truancy became easy:

School 2, group 2:

S2: There are actually a lot of groups that do not do anything in those lessons (. . .) I've seen many groups who just go home, or something.

S3: They say they're going on a hike, but they just hike to the bus stop (laughter).

(. . .)

S1: Yes, I think few actually attend the BH lessons

Interviewer: Have you or anyone else said something about this to the teachers?

S1: No

S2: Snitches get stitches, (laughter)

The participants' interest and motivation also influenced fidelity, in particular regarding the planning process and use of the activity contract. Though participants spent time planning and writing the activity contract, adherence to the activity contract varied within and between schools. In Schools 1, 2 and 4, most groups pursued only one activity and the lessons were more formal than in School 3. In School 3, some groups wrote "various activities" on their activity contract, while others decided on the spot what to do in a given lesson. Some groups also decided to play together, regardless of what they had written in their activity plans. Although the participants initiated this adaptation, an enabling influencing factor was the provider, who allowed it. The adaptation did not seem to have a negative influence on participation or efforts:

School 3, group 2:

S2: Leadership?

S1: No, there's no leadership

S2: We're pretty much, like, we lead each other

S1: We're all leaders, so one of us might say "let's play soccer today", then we play soccer (. . .) we just run in, get a ball, and start playing.

Conflicts between students emerged as a factor influencing fidelity and was mainly reported in School 4, where they struggled with conflicts within female groups. According to the teachers, DWBH did not cause the conflicts; rather, DWBH helped expose pre-existing hidden conflicts. The conflicts forced the teachers to intervene, because the students could not or would not try to solve the conflicts themselves. This resulted in reduced participation and altered group compositions and a lack of fidelity to groups' conflict resolution plans:

School 4, teachers:

T3: We've definitely steered some students here and there (laughter).

(. . .)

T2: We're still working on the worst conflicts. But conflicts have been solved by some changing groups.

Provider characteristics

The providers, that is, the teachers, emerged as important influencers of all aspects of implementation. In School 1, one of the two teachers providing DWBH was on long-term sick leave, leaving the second teacher alone with two classes. He expressed a limited ability to supervise all the students because he was mostly compelled to be present in the gymnasium. Sometimes, a substitute teacher was also present, which allowed the main teacher to visit activity groups outside the gymnasium. The teacher maintained that although he could handle being the sole supervisor, two teachers were necessary to ensure all students felt they

were seen and to act as a mediator in case of a conflict between students. Regardless, the remaining teacher's actions and status among the students were positively influencing factors in School 1:

> School 1, group 3:
>
> S2: He is the kind of teacher that you have a good relationship with, and you're not afraid to talk to him or anything. He is a really good teacher (. . .) Sometimes he plays soccer with us.
>
> S1: It's fun.

As the quote suggests, the students in School 1 appreciated the teacher's participation in their activity. Similar positive responses occurred in School 3, where the teacher also participated with students on occasion. There were no guidelines from project management regarding teacher participation in the students' activities, but the teacher in School 1 felt that it had a positive influence on his relationship with the students. The students from Schools 1, 3 and 4 spoke of their teachers in either positive or neutral terms. The students from School 2, on the other hand, were more critical of their teachers. For instance, they could not remember what they had written in their activity contract, which they blamed on their teacher. Allegedly, he had collected the contracts when they were completed and the students had not seen them since. One of the groups that had activities off school grounds went even further in its negative description of the teachers:

> School 2, group 3:
>
> S2: They definitely do not care about the project. I met a teacher down at the mall after school when I was supposed to be in the Be Happy lesson, and she was like "hi, should not you be in Be Happy?". I was like "yeah", and she just said "okay". Like, the teachers at this school are so bad.

Discussion

The main findings show large differences between schools regarding how DWBH was implemented and how various factors influenced the implementation. Schools 1 and 4 made minor adaptations in the way DWBH was organized, and these were positively received by the students. Intervention characteristics, spacious facilities, scheduling and participant and provider characteristics positively influenced all aspects of implementation. School 2 made major adaptations to how DWBH was scheduled, which reduced both *i* responsiveness and fidelity. Additionally, limited facilities and participant and provider characteristics negatively impacted fidelity, quality and dose received. School 2 was the only school where the intervention was negatively received. School 3 made one major adaptation in how DWBH was organized, and it was poorly received by the students. The intervention itself was otherwise positively received. Limited facilities and scheduling negatively impacted fidelity. Intervention, participant and provider characteristics positively influenced responsiveness, quality and perhaps also dose received.

Reasons for and consequences of adaptations

A common adaptation in Schools 1, 3 and 4 was that the teachers decided to have two identical BH lessons rather than DW and BH. The adaptation was made because the teachers wanted to and because they thought it fit with the purpose of the intervention. Students and teachers agreed that it was a positive adaptation. The adaptations in School 2, however, were mostly made because of contextual limitations. For instance, two periods were added to the schedule instead of one, because all grades and classes followed a fixed schedule, so they could not reorganize the schedule for only some of the students. The adaptation reduced

students' leisure time, which caused a negative response before the intervention had even started. Contextual limitations were the reasons for the main adaptation in School 3 as well and also caused a negative response among the affected students. Initially, limited gymnasium space compelled the teacher to carry out BH in separate classes. However, one of the classes had access to the gymnasium during only one of the two weekly lessons, because it was being used by another PE class. These adaptations and their respective reactions can be elucidated by Moore *et al.* (2013), who claim that adaptations can be either positive, neutral or negative and either logistical or philosophical. An intervention can be adapted to fit the context and positively influence implementation (Durlak and Dupre, 2008; Berkel *et al.*, 2011), which is what happened in Schools 1 and 4, and to some extent in School 3, where they made positive adaptations for philosophical reasons. In Schools 2 and 3, however, negative adaptations for logistical reasons negatively impacted implementation. The findings involving adaptations indicate that schools that were likely (because of their preconditions) to succeed in implementing DWBH anyway, made the positive adaptations. Conversely, the schools that were less likely (because of their preconditions) to succeed in implementing DWBH made the negative adaptations.

Dose is not enough

The findings indicated frequent truancy among many students in School 2, caused by negative responsiveness, in combination with poor facilities, scheduling BH to the last lesson of the day and limited teacher supervision. In contrast, positive responsiveness and student interest positively influenced participation in Schools 1, 3 and 4. This concurs with the model designed by Berkel *et al.* (2011), which proposed that dose received is partly determined by responsiveness. In the review by Naylor *et al.* (2015), a similar factor to responsiveness, "student characteristics, engagement and motivation," was presented as one of 22 factors influencing dose delivered/received. Although dose received is not an aspect of implementation as defined by Durlak and Dupre (2008), it was included in the review by Naylor *et al.* (2015), as being one of the most frequently used measures of implementation for school-based PA interventions, along with dose delivered. Dose delivered and dose received provide important information about amount; however, the present study revealed details about the specific components of the intervention that were actually executed, how well they were executed, how students and teachers experienced them and whether these aspects interacted in some way. The present findings suggest that a single focus on dose delivered/received may be a somewhat limited view of implementation that says little about how suitable the intervention is in any school context. This coincides with the findings of Tarp *et al.* (2016), who conducted a school-based PA intervention cluster RCT and found no effect. According to objective measurements, they could not deliver their target PA dose, but they could not explain why. They therefore recommended that qualitative data on implementation be included in future studies to improve the ability to explain results.

The present results also raise concerns about quantitatively measuring dose delivered without comparing it to dose received. Delivery alone tells us nothing about motivation, interest or actual participation, which has more impact on the intervention outcomes than delivery (Khanal *et al.*, 2019; Roth, 1985; Durlak and Dupre, 2008). Dose delivered is easy to measure and can be an accurate depiction of the amount provided. However, if dose delivered is the only measure of implementation, researchers may erroneously assume a successful implementation on the grounds of high delivery rates, while low levels of fidelity, quality and responsiveness remain unobserved. On the other hand, and despite plenty of qualitative information, the present results do not tell us how important fidelity, quality and responsiveness to DWBH are for achieving the expected outcomes for physical fitness, mental health, academic achievement or learning environment. The results only indicate that the majority of students and teachers found DWBH to be relevant and enjoyable when (1)

DWBH had few or no consequences for students' leisure time; (2) DWBH was executed with adequate facilities; (3) DWBH was provided by teachers who were present and cared about what the students did; and (4) adaptations were perceived positively and did not negatively impact (1), (2) or (3).

Context and suitability

The present results reveal the complexity of the school context, how schools can differ and how differently schools can carry out a complex and demanding intervention. For instance, facilities were an important factor that differentiated the schools and influenced how DWBH, from the beginning, was adapted, differently, in each school. On top of this, a dynamic interaction between intervention components, teachers and students determined how the implementation process developed. Furthermore, this development occurred differently on multiple levels: the school level (e.g. different facilities, scheduling and teachers), class level (e.g. the scheduling problem in School 3, where one of the participating classes did not have facilities during DW), activity-group level (e.g. few groups within a school pursued the same activities) and student level (e.g. students lost interest and changed groups). Moore *et al.* (2019) argued that introducing a complex intervention in a complex system poses an almost infinite number of uncertainties, which no evaluation is able to address completely. That may be the case, but the process evaluation enables us to address the suitability of DWBH in Norwegian lower secondary schools, which is required in order to say anything about the intervention's generalizability (Bonell *et al.*, 2006). The somewhat limited implementation that occurred in two out of four schools indicates that DWBH may not have been suitable for these schools. Assuming that other lower secondary schools are as varied as our sample, a strict DWBH program may not be generalizable. However, with the knowledge of the factors that influenced implementation of DWBH, it might be possible to adapt the intervention in a way that fits all contexts, without compromising its purpose. Furthermore, the main negatively influencing factors in Schools 2 and 3 were perhaps not caused by an unsuitable intervention, but by the schools being unable to introduce any program on short notice. Thus, the generalizability of DWBH as a sustainable way to increase PA in lower secondary schools remains uncertain, although we must underscore that having suitable facilities might be the most important precondition.

Design challenges

RCTs are regarded as the gold standard for evaluating effectiveness of public health interventions (Moore *et al.*, 2015; Victora *et al.*, 2004), and the use of RCTs has been contested (Byrne, 2013) and defended (Hawe *et al.*, 2004). Causal inferences from an RCT are based on outcome comparisons between the intervention group and the control group (Rubin, 1974) and depend on randomization to eliminate differences in observed or unobserved variables between the groups. However, the present results indicate that at least three variables (frame factors, participant characteristics and provider characteristics) varied greatly between schools and may cause systematic differences between intervention group and control group. To avoid these differences, a matched pairs design, based on key frame factors such as facilities, could have been a viable option (Stuart, 2010). However, in our cluster RCT, the problem was not necessarily differences between intervention group and control group, but large differences within the intervention group. The differences caused DWBH to be implemented differently to the extent that students in different schools received different interventions. DWBH outcomes were most likely influenced by these differences (Durlak and Dupre, 2008) and randomization, unfortunately, is not a solution to the problem. It is possible that if we had certain predetermined inclusion criteria, such as facilities, we would only have recruited schools that were able to accommodate the intervention. The results might then

have shown fewer implementation differences between the schools. Although this would reduce the representativeness of the included schools, we cannot expect an outcome to change if the intervention school is unable to accommodate the intervention. The variation in facilities may have been the single aspect that mattered the most for the implementation quality, further underscoring the importance of including the aspect in process evaluations. In future cluster RCTs for school-based PA interventions, it is therefore important that researchers ask themselves "what will the intervention require from the school, if it is to be implemented with high quality?" and recruit schools accordingly.

When interpreting the results from a complex cluster-RCT intervention, the results from a process evaluation represent an invaluable tool, and the evaluation should always be conducted whenever there may be variability in the implementation process (Craig *et al.*, 2008; Oakley *et al.*, 2006). Traditionally, the RCT attempts to answer the question "what works?" Combining the RCT with a process evaluation helps us answer "why things work?" (Deaton and Cartwright, 2018) and "under what circumstances?" (Bonell *et al.*, 2012). Answering these questions is essential to designing school-based PA interventions that are feasible for large-scale implementation.

Strengths and limitations

By combining qualitative information on fidelity, adaptation, quality, responsiveness and dose received with quantitative information on dose delivered, the present study provides detailed information regarding the implementation of a school-based PA intervention, compared to previous research (Naylor *et al.*, 2015; Watson *et al.*, 2017; Daly-Smith *et al.*, 2018). Previously conducted process evaluations have highlighted the use of quantitative and qualitative methods as important strengths (De Meij *et al.*, 2013; Burges Watson *et al.*, 2016). However, this study has several limitations. Limited time and resources compelled us to conduct interviews in only four out of nine intervention schools and at only one point in time, rather than multiple times, as recommended by the literature (Moore *et al.*, 2015). As implementations can change over time (Dusenbury *et al.*, 2003), it is important to note that the present findings might not represent implementation throughout the intervention period. Moreover, considering that the qualitative data collection spanned over two months, the implementation process in the first school evaluated might have been at a different stage than in the last school evaluated. Finally, the authors of this paper were stakeholders in the cluster RCT evaluating DWBH and may have an interest in portraying it positively. This may have influenced how we conducted the process evaluation, interpreted and reported the results.

Conclusion

This process evaluation showed that two out of four qualitatively examined schools delivered the intervention with high fidelity, quality, dose delivered and dose received, while obtaining positive responsiveness from participants and providers. The other two schools had major adaptations and limitations and delivered the intervention with varying fidelity, quality, dose delivered and dose received. Frame factors, intervention characteristics, participant characteristics and provider characteristics influenced implementation, and differences between schools may impact the intervention outcomes. Positive adaptations were made in schools that were likely to succeed anyway, based on their preconditions, while negative adaptations were made in schools that, based on their preconditions, were less likely to succeed. The results indicate that adequate facilities and scheduling that did not affect participants' leisure time were important to ensure that the intervention was positively received. Negative responsiveness negatively influenced dose received. Future school-based PA interventions should be designed to generate positive response, perhaps by organizing student-led lessons. However, if responses are negative in certain schools, providers,

researchers and students should cooperate to adapt the intervention in order to make it more relevant and suitable. Careful monitoring of multiple aspects of implementation is key to be able to act upon such responses. We therefore recommend that qualitative process evaluations are conducted on future trials involving school-based PA interventions.

References

Agans, J.P., Sävfenbom, R., Davis, J.L., Bowers, E.P. and Lerner, R.M. (2013), "Positive movement experiences: approaching the study of athletic participation, exercise, and leisure activity through relational developmental systems theory and the concept of embodiment", *Advances in Child Development and Behavior,* Vol. 45, pp. 261-286.

Barriball, K.L. and While, A. (1994), "Collecting data using a semi-structured interview: a discussion paper", *Journal of Advanced Nursing-Institutional Subscription,* Vol. 19 No. 2, pp. 328-335.

Berkel, C., Mauricio, A.M., Schoenfelder, E. and Sandler, I.N. (2011), "Putting the pieces together: an integrated model of program implementation", *Prevention Science,* Vol. 12 No. 1, pp. 23-33.

Bonell, C., Oakley, A., Hargreaves, J., Strange, V. and Rees, R. (2006), "Assessment of generalisability in trials of health interventions: suggested framework and systematic review", *British Medical Journal,* Vol. 333 No. 7563, pp. 346-349.

Bonell, C., Fletcher, A., Morton, M., Lorenc, T. and Moore, L. (2012), "Realist randomised controlled trials: a new approach to evaluating complex public health interventions", *Social Science and Medicine,* Vol. 75 No. 12, pp. 2299-2306.

Burges Watson, D., Adams, J., Azevedo, L.B. and Haighton, C. (2016), "Promoting physical activity with a school-based dance mat exergaming intervention: qualitative findings from a natural experiment", *BMC Public Health,* Vol. 16 No. 1, p. 609.

Byrne, D. (2013), "Evaluating complex social interventions in a complex world", *Evaluation,* Vol. 19 No. 3, pp. 217-228.

Clarke, A.M. (2010), "Context matters in programme implementation", *Health Education,* Vol. 110 No. 4, pp. 273-293.

Craig, P., Dieppe, P., Macintyre, S., Michie, S., Nazareth, I. and Petticrew, M. (2008), "Developing and evaluating complex interventions: the new Medical Research Council guidance", *BMJ,* Vol. 337, p. 1655.

Daly-Smith, A.J., Zwolinsky, S., Mckenna, J., Tomporowski, P.D., Defeyter, M.A. and Manley, A. (2018), "Systematic review of acute physically active learning and classroom movement breaks on children's physical activity, cognition, academic performance and classroom behaviour: understanding critical design features", *BMJ Open Sport and Exercise Medicine,* Vol. 4 No. 1, 000341.

De Meij, J.S.B., Van Der Wal, M.F., Van Mechelen, W. and Chinapaw, M.J.M. (2013), "A mixed methods process evaluation of the implementation of JUMP-in, a multilevel school-based intervention aimed at physical activity promotion", *Health Promotion Practice,* Vol. 14 No. 5, pp. 777-790.

Deaton, A. and Cartwright, N. (2018), "Understanding and misunderstanding randomized controlled trials", *Social Science and Medicine,* Vol. 210, pp. 2-21.

Demetriou, Y. and Honer, O. (2012), "Physical activity interventions in the school setting: a systematic review", *Psychology of Sport and Exercise,* Vol. 13 No. 2, pp. 186-196.

Dobson, D. and Cook, T.J. (1980), "Avoiding type III error in program evaluation: results from a field experiment", *Evaluation and Program Planning,* Vol. 3 No. 4, pp. 269-276.

Durlak, J.A. (2016), "Programme implementation in social and emotional learning: basic issues and research findings", *Cambridge Journal of Education,* Vol. 46 No. 3, pp. 333-345.

Durlak, J.A. and Dupre, E.P. (2008), "Implementation matters: a review of research on the influence of implementation on program outcomes and the factors affecting implementation", *American Journal of Community Psychology,* Vol. 41 Nos 3-4, pp. 327-350.

Dusenbury, L., Brannigan, R., Falco, M. and Hansen, W.B. (2003), "A review of research on fidelity of implementation: implications for drug abuse prevention in school settings", *Health Education Research,* Vol. 18 No. 2, pp. 237-256.

Frey, J.H. and Fontana, A. (1991), "The group interview in social research", *The Social Science Journal,* Vol. 28 No. 2, pp. 175-187.

Gale, N.K., Heath, G., Cameron, E., Rashid, S. and Redwood, S. (2013), "Using the framework method for the analysis of qualitative data in multi-disciplinary health research", *BMC Medical Research Methodology,* Vol. 13 No. 1, p. 117.

Hawe, P., Shiell, A. and Riley, T. (2004), "Complex interventions: how "out of control" can a randomised controlled trial be?", *BMJ,* Vol. 328 No. 7455, pp. 1561-1563.

Hawe, P., Shiell, A. and Riley, T. (2009), "Theorising interventions as events in systems", *American Journal of Community Psychology,* Vol. 43 Nos 3-4, pp. 267-276.

Humphrey, N., Barlow, A. and Lendrum, A. (2017), "Quality matters: implementation moderates student outcomes in the PATHS curriculum", *Prevention Science*, Vol. 19, pp. 197-208.

Hynynen, S.T., Van Stralen, M.M., Sniehotta, F.F., Araújo-Soares, V., Hardeman, W., Chinapaw, M.J.M., Vasankari, T. and Hankonen, N. (2016), "A systematic review of school-based interventions targeting physical activity and sedentary behaviour among older adolescents", *International Review of Sport and Exercise Psychology,* Vol. 9 No. 1, pp. 22-44.

Keshavarz, N., Nutbeam, D., Rowling, L. and Khavarpour, F. (2010), "Schools as social complex adaptive systems: a new way to understand the challenges of introducing the health promoting schools concept", *Social Science and Medicine,* Vol. 70 No. 10, pp. 1467-1474.

Khanal, S., Choi, L., Innes-Hughes, C. and Rissel, C. (2019), "Dose response relationship between program attendance and children's outcomes in a community based weight management program for children and their families", *BMC Public Health,* Vol. 19 No. 1, p. 716.

Kolle, E., Solberg, R.B., Grydeland, M., Säfvenbom, R., Anderssen, S.A., Ekelund, U., Berntsen, S., Resaland, G.K., Dyrstad, S.M. and Steene-Johannessen, J. (2020), "Evaluating the effect of increased physical activity in lower secondary school adolescents: the study protocol for the school in motion (ScIM) study – a cluster-randomized controlled trial", Mansucript submitted for publication.

Kriemler, S., Meyer, U., Martin, E., Van Sluijs, E.M.F., Andersen, L.B. and Martin, B.W. (2011), "Effect of school-based interventions on physical activity and fitness in children and adolescents: a review of reviews and systematic update", *British Journal of Sports Medicine,* Vol. 45 No. 11, pp. 923-930.

Lerner, R.M. (2015), "Promoting positive human development and social justice: integrating theory, research and application in contemporary developmental science", *International Journal of Psychology,* Vol. 50 No. 3, pp. 165-173.

Lerner, R.M., Hershberg, R., Hilliard, L. and Johnson, S. (2015), "Concepts and theories in human development", in Bornstein, M.H. and Lamb, M.E. (Eds), *Developmental Science: An Advanced Textbook*, Psychology Press, New York, pp. 3-42.

Linnan, L. and Steckler, A. (2002), "Process evaluation for public health interventions and research", in Steckler, A. and Linnan, L. (Eds), *Process Evaluation for Public Health Interventions and Research*, Jossey-Bass, San Francisco, pp. 1-23.

Lubans, D.R., Smith, J.J., Morgan, P.J., Beauchamp, M.R., Miller, A., Lonsdale, C., Parker, P. and Dally, K. (2016), "Mediators of psychological well-being in adolescent boys", *Journal of Adolescent Health,* Vol. 58 No. 2, pp. 230-236.

Manzano, A. (2016), "The craft of interviewing in realist evaluation", *Evaluation,* Vol. 22 No. 3, pp. 342-360.

Mays, N. and Pope, C. (1995), "Qualitative Research: observational methods in health care settings", *BMJ,* Vol. 311 No. 6998, pp. 182-184.

Moore, J., Bumbarger, B. and Cooper, B. (2013), "Examining adaptations of evidence-based programs in natural contexts", *Journal of Primary Prevention,* Vol. 34 No. 3, pp. 147-161.

Moore, G.F., Audrey, S., Barker, M., Bond, L., Bonell, C., Hardeman, W., Moore, L., O'cathain, A., Tinati, T., Wight, D. and Baird, J. (2015), "Process evaluation of complex interventions: medical Research Council guidance", *BMJ British Medical Journal,* Vol. 350, p. 7.

Moore, G.F., Evans, R.E., Hawkins, J., Littlecott, H., Melendez-Torres, G.J., Bonell, C. and Murphy, S. (2019), "From complex social interventions to interventions in complex social systems: future directions and unresolved questions for intervention development and evaluation", *Evaluation,* Vol. 25 No. 1, pp. 23-45.

Naylor, P.-J., Nettlefold, L., Race, D., Hoy, C., Ashe, M.C., Wharf Higgins, J. and Mckay, H.A. (2015), "Implementation of school based physical activity interventions: a systematic review", *Preventive Medicine,* Vol. 72, pp. 95-115.

Norris, E., Shelton, N., Dunsmuir, S., Duke-Williams, O. and Stamatakis, E. (2015), "Physically active lessons as physical activity and educational interventions: a systematic review of methods and results", *Preventive Medicine,* Vol. 72, pp. 116-125.

Oakley, A., Strange, V., Bonell, C., Allen, E. and Stephenson, J. (2006), "Process evaluation in randomised controlled trials of complex interventions", *British Medical Journal,* Vol. 332 No. 7538, pp. 413-416.

Revicki, D.A. and Frank, L. (1999), "Pharmacoeconomic evaluation in the real world", *PharmacoEconomics,* Vol. 15 No. 5, pp. 423-434.

Rickles, D. (2009), "Causality in complex interventions", *Medicine, Healthcare and Philosophy,* Vol. 12 No. 1, pp. 77-90.

Robinson, O.C. (2014), "Sampling in interview-based qualitative research: a theoretical and practical guide", *Qualitative Research in Psychology,* Vol. 11 No. 1, pp. 25-41.

Roth, H.J. (1985), "Relationship between attendance at a parents anonymous adult program and children's behavior at the parents anonymous child care program", *Children and Youth Services Review,* Vol. 7 No. 1, pp. 39-43.

Rubin, D.B. (1974), "Estimating causal effects of treatments in randomized and nonrandomized studies", *Journal of Educational Psychology,* Vol. 66 No. 5, pp. 688-701.

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. (2018), "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,* Vol. 166 No. 1, pp. 49-55.

Smith, J., Beauchamp, M., Faulkner, G., Morgan, P., Kennedy, S. and Lubans, D.R. (2018), "Intervention effects and mediators of well-being in a school-based physical activity program for adolescents: the 'Resistance Training for Teens' cluster RCT", *Journal of Mental Health and Physical Activity,* Vol. 15, pp. 88-94.

Spencer, L. and Ritchie, J. (2002), "Qualitative data analysis for applied policy research", in Bryman, A. and Burgess, R.G. (Eds), *Analyzing Qualitative Data*, Routledge, London, pp. 187-208.

Stuart, E.A. (2010), "Matching methods for causal inference: a review and a look forward", *Statistical Science: A Review Journal of the Institute of Mathematical Statistics,* Vol. 25 No. 1, pp. 1-21.

Tarp, J., Domazet, S.L., Froberg, K., Hillman, C.H., Andersen, L.B. and Bugge, A. (2016), "Effectiveness of a school-based physical activity intervention on cognitive performance in Danish adolescents: LCoMotion—learning, cognition and motion – a cluster randomized controlled trial", *PloS One,* Vol. 11 No. 6, e0158087.

Victora, C.G., Habicht, J.-P. and Bryce, J. (2004), "Evidence-based public health: moving beyond randomized trials", *American Journal of Public Health,* Vol. 94 No. 3, pp. 400-405.

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*, Vol. 14, article 114.

Corresponding author
Andreas Avitsland can be contacted at: andreas.avitsland@uis.no

Paper 3

Effects of school-based physical activity interventions on mental health in adolescents: The School in Motion cluster randomized controlled trial

Mental Health and Physical Activity, in press (doi online journal pre-proof: https://doi.org/10.1016/j.mhpa.2020.100348)

Andreas Åvitsland[1], Eva Leibinger[1], Geir Kåre Resaland[2], Runar Barstad Solberg[3], Elin Kolle[3], Sindre M. Dyrstad[1, 4]

[1]*Department of Teacher Education and Sports Sciences, University of Stavanger, Norway*
[2]*Center for physically active learning, Western Norway University of Applied Sciences, Norway*
[3]*Department of Sports Medicine, Norwegian School of Sports Sciences, Norway*
[4]*Department of Public Health, University of Stavanger, Norway*

Corresponding author: Andreas Åvitsland, Andreas.avitsland@uis.no

Declarations of interest: The authors declare that they have no competing interests.

Acknowledgments: Thanks to all schools, principals, teachers, students and parents, who agreed to participate.

Abstract

Purpose: To investigate the effects of two school-based physical activity interventions on mental health in Norwegian adolescents.

Methods: Students from 29 lower secondary schools in Norway (n = 2084; 14-15 years; 49% female) were cluster-randomized into either a control group or one of two intervention groups (M1 and M2). Two interventions based on different theoretical frameworks aimed to increase physical activity in school by approximately 120 minutes per week, throughout a 29-week intervention period. M1 consisted of 30 minutes physically active learning, 30 minutes physical activity and one 60 minutes physical education lesson. M2 consisted of one physical education lesson and one physical activity lesson, both focusing on facilitating students' interest, responsibility and social relationships. The self-report version of the Strengths and Difficulties Questionnaire was used to assess mental health. Physical activity was measured by accelerometry. Linear mixed effects models were used to examine the effects of the interventions.

Results: No effects were found for the overall study population. Interaction effects warranted subgroup analyses: M1 showed favorable results in the subgroup with the highest levels of psychological difficulties at baseline (b = -2.9; -5.73 to -0.07; p = .045) and in the immigrant subgroup (b = -1.6; -3.53 to 0.27; p = .093). M2 showed favorable results in the immigrant subgroup (b = -2.1; -4.36 to 0.21; p = .075).

Conclusions: The two interventions did not improve mental health in the full study population. However, results indicated beneficial effects among immigrants and those with poor mental health at baseline. More research is needed due to missing values and the results should therefore be interpreted with caution.

Keywords: Physical activity, mental health, school-based intervention, cluster randomized controlled trial

1. Introduction

The leading causes for disability among children and adolescents worldwide are mental- and substance use disorders (Erskine et al., 2015). In Norway, 10% of boys and 28% of girls graduating from lower secondary school report mental health problems (Bakken, 2019). Furthermore, from 2014 to 2019, the amount of Norwegian adolescent boys and girls reporting mental health problems has increased, respectively, from 8% to 10% and 21% to 27% (Bakken, 2015; Bakken 2019). Similar increases in adolescent mental health problems have also been found internationally (Collishaw, 2015). For adolescents, mental disorders such as depression have been associated with poor academic achievement, low school attendance and alcohol and drug use (Fröjd et al., 2008; Glied & Pine, 2002). Adolescents who experience mental disorders are also more likely to develop similar or more severe conditions as adults (Kessler et al., 2007). Adolescent mental health therefore also represents an economic challenge as mental disorders topped the list of the costliest conditions in Norway in 2013 (Kinge, Sælensminde, Dieleman, Vollset, & Norheim, 2017). The above mentioned studies provide a solid foundation to establish feasible methods for the purpose of preventing these problems.

Physical activity (PA) as a treatment against mental health problems in clinical adolescent populations has been researched extensively, and two recent reviews concluded that a large variety of PA interventions and moderate to high intensity aerobic exercise were likely to have a positive effect on depression (Bailey, Hetrick, Rosenbaum, Purcell, & Parker, 2017; Biddle, Ciaccioni, Thomas, & Vergeer, 2019). Studies examining the potential effect of PA on internalizing and externalizing problems in generally healthy adolescent community populations show mixed or weak results, indicating a potential ceiling effect due to a smaller potential for improvement (Spruit, Assink, van Vugt, van der Put, & Stams, 2016). However, the data is inconclusive and further research in community populations (Biddle et al., 2019), for instance, among adolescent school students, is warranted. Schools are optimal arenas for intervening, as researchers can reach students equally across sex, SES and ethnicity. The World Health Organization (WHO) has recommended that schools take part in promoting PA, for the purpose of raising children's and adolescents' PA levels and improving health (WHO, 2010). Since students spend a significant portion of the day at school, there is an opportunity to facilitate increased PA during school hours (Hills, Dengel, & Lubans, 2015). Potentially, this would make an impact on Norwegian 15-year-olds' PA levels, as only half of this demographic meet the recommended 60 minutes of daily moderate to vigorous PA (Dalene et al., 2018).

Additionally, concerns that cognition or academic achievement would be negatively affected by increasing school-based PA at the expense of theoretical subjects are not supported by research (Donnelly et al., 2016; A. S. Singh et al., 2019).

The few studies examining how school-based PA interventions affect mental health in children and adolescents show mixed results. Bonhauser et al. (2005) Casey et al. (2014) and Lubans et al. (2016) examined low-SES populations and showed, respectively, a decrease in anxiety and an increase in self-esteem, improvements in health-related quality of life for adolescent girls, and improved well-being for boys. Christiansen et al. (2018), Smith et al. (2018) and Eather, Morgan, and Lubans (2016) examined generally healthy community populations and found no overall effect. However, Smith et al. (2018) found a tendency toward an effect on self-esteem in the overweight/obese subgroup. Eather et al. (2016) found that the subgroup with the most psychological difficulties at baseline showed beneficial effects on self-esteem, perceived body fat, perceived appearance and physical self-concept. Lastly, Christiansen et al. (2018) found improvements in self-worth for students who did not participate in leisure-time sports. These findings suggest that school-based PA interventions are unlikely to elicit detectable effects on mental health outcomes in generally healthy adolescent populations. However, the findings also suggest that certain subgroups may benefit more from a school-based PA intervention than the average population. This substantiates findings of Cerin (2010), which indicated that PA effects can be heterogeneous, and therefore warrants investigation of relevant subgroups. In this context, relevant subgroups have been shown to display lower PA levels and poorer mental health than the population outside of the subgroups. This is the case among immigrants (Abebe, Lien, & Hjelde, 2014; Sagatun, Kolle, Anderssen, Thoresen, & Søgaard, 2008; Singh, Yu, Siahpush, & Kogan, 2008), low socioeconomic status (SES) populations (Bøe, Øverland, Lundervold, & Hysing, 2012; Heelan et al., 2010), girls (Bakken, 2019; Dalene et al., 2018) and poor mental health populations (Pinto Pereira, Geoffroy & Power, 2014). Although school-based PA interventions have been shown to be effective on mental health in low-SES groups (Casey et al., 2014; Lubans et al., 2016) and poor mental health groups (Eather et al., 2016), the knowledge base is limited. Furthermore, PA has been shown to be beneficial for mental health among immigrants (Siddiqui, Lindblad, & Bennet, 2014); however, to the authors' knowledge, no studies have examined the association between PA and mental health among immigrant adolescents. Although the use of subgroup analyses is debated (Sun et al., 2012; Wang & Ware, 2013), Biddle et al. (2019) recommended that future research should focus on

the potentially different effects between sexes and between those with different mental health conditions.

The primary aim of the present paper was to assess the effect of two school-based PA interventions on adolescents' mental health. We hypothesized heterogeneous effects, so a secondary aim was to analyze subgroups. A third aim was to examine the subscales of the mental health instrument to establish which aspects of mental health a potential effect could be attributed to.

2. Method

2.1 Design and participants

This paper presents data from the School in Motion study. Briefly, School in Motion was a multicenter study, designed as a cluster randomized controlled trial (RCT) involving four test centers in Norway. The primary aim of the study was to assess whether two different school-based PA interventions affected PA levels. The secondary aims were to assess the effects on physical fitness, mental health, learning environment and academic achievement. Thirty lower secondary schools accepted the invitation to participate and were randomized into three groups: two intervention groups and one control group. A neutral third party was responsible for the randomization process, after which, one of the schools withdrew from the study. Students attending ninth grade during the intervention period were invited to participate, and we obtained informed parental consent from 76% of the eligible students (n = 2084). The intervention period was 29 weeks. The study is registered in ClinicalTrials.gov ID nr: NCT03817047. Fig. 1 shows the participant flow from enrollment to post-testing.

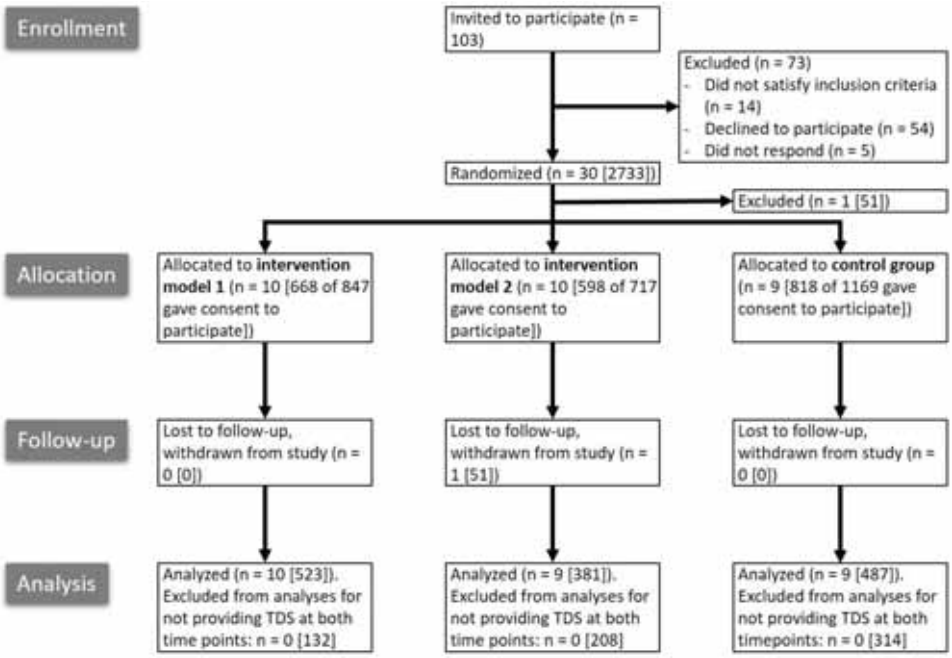


Fig. 1. The flow diagram shows the flow of schools and participants through the School in Motion-study. Numbers represent schools [children]. TDS = Total difficulties score.

2.2 Interventions

Intervention model 1 (M1), named "Active learning", consisted of weekly physically active academic lessons (30 min/week), PA not connected to a curriculum (30 min/week), and one additional physical education (PE) lesson (45-60 min/week). The purpose of M1 was to increase PA levels and assess the feasibility of incorporating PA into theoretical subjects in lower secondary school. We encouraged the schools to incorporate PA into all theoretical subjects, but they were ultimately in charge of which subjects they wanted to be included in the project. The intervention draws on three theoretical perspectives: Physical Literacy (Whitehead, 2010), Self-efficacy (Bandura, 1982) and Basic Psychological Needs Theory (Ryan & Deci, 2002). In short, the intervention is theorized to increase students' motivation for PA, to let students' physical learning influence other types of learning and to increase self-efficacy. In turn, increased self-efficacy is associated with mental health outcomes, such as anxiety and depression (Muris, 2002). Intervention model 2 (M2), named "Don't worry, be happy", consisted of one additional PE lesson (45-60 min/week) and one additional PA lesson (45-60 min/week). The purpose of M2 was not only the PA dose itself, but also to encourage students

to pursue activities of their own interest in groups they formed themselves. They were allowed to choose what they wanted to do and small "activity groups" were formed based on students' choices. The intervention draws on three theoretical perspectives: Positive Youth Development (Lerner & Lerner, 2013), Relational Developmental Systems (Lerner, Hershberg, Hillard & Johnson, 2015) and Positive Movement Experiences (Agans, Sävfenbom, Davis, Bowers, & Lerner, 2013). In short, the intervention is theorized to facilitate that students develop social relationships and experience positive emotions while participating in activities that are meaningful to them. In turn, the intervention is thought to increase participants' motivation for PA and influence mental health through a psychosocial mechanism. Although activities in M2 were mostly student-led, both interventions were formally delivered by the teachers. Extended descriptions of the design and implementation of the M2-intervention has been published elsewhere (Åvitsland, Ohna, et al., 2020).

Teachers from both intervention groups participated in a workshop where they received instructions and training. In addition, halfway through the intervention period, another workshop was organized, in which teachers from different schools could discuss their progress and solutions with each other and the researchers. Considering the design of M2, teachers providing this intervention did not require extensive training. The teachers in M1, however, had access to an online "tool-kit" containing various suggestions for physically active academic lessons. Throughout the intervention period, researchers from each test center kept in touch regularly with schools via emails and visited the schools at least twice per semester.

2.3 Measurements

Mental health

To measure mental health, the participants completed the Strengths and Difficulties Questionnaire (SDQ; Goodman, 1997) at baseline (T1, May-August 2017) and at follow-up, (T2, April-June 2018). The SDQ is a self-report questionnaire consisting of 25 items, divided into five subscales, each containing five items. The main outcome variable is psychological difficulties, expressed by the total difficulties score (TDS), which is made up of four subscales: 1) emotional problems (worry, unhappy, nervous, scared), 2) conduct problems (temper tantrums, fights, lies, steals), 3) hyperactivity (restless, fidgety, distracted) and 4) peer problems (solitary, not liked, bullied). The subscales score from 0-10 and TDS ranges from 0-40. A higher score signifies increased psychological difficulties. In large populations, TDS can detect

changes in psychopathology on each point of the scale, and therefore it can be seen as a dimensional measure that can indicate a general mental health state in children and adolescents (Goodman & Goodman, 2009). The TDS scale can be divided into three levels, in order to identify the risk of mental disorders: “normal” (0-15), “borderline” (16-19) and “abnormal” (20-40), which is characterized as being at risk of developing mental disorders (Goodman, Ford, Simmons, Gatward, & Meltzer, 2000). The psychometric properties of the SDQ have been validated in many countries, including Norway (Van Roy, Veenstra, & Clench-Aas, 2008).

Subgroups

To allow for the examination of interaction effects and subgroup analyses, participants self-reported their sex and status as either immigrant or non-immigrant. Immigrant status was determined by foreign or native birthplace (born in Norway, yes/no). We used national registries to obtain SES, which is expressed as the parent with the highest education level (Erola, Jalonen, & Lehti, 2016). SES was divided into four subgroups: 1) lower secondary school or less, 2) upper secondary school, 3) less than four years university education, and 4) four years or more university education. We created baseline TDS subgroups based on the predetermined cutoffs defining “normal”, “borderline” and “abnormal” TDS. This paper adheres to the three most critical criteria used to assess credibility of subgroup effects by Sun et al. (2012): subgroup variables must be assessed at baseline, subgroup hypotheses must be specified ahead of analyses and there must be an interaction effect.

Physical activity

PA was objectively measured with Actigraph accelerometers, models GT3X and GT3X+ (Actigraph, LLC, Pensacola, Florida, USA). Measurements were carried out over seven consecutive days, and participants were instructed to wear the accelerometer on their right hip, and to take it off during sleep, or when in contact with water. The accelerometers were initialized to start recording at 06:00 the day after the participants started wearing them. We excluded data recorded between 00:00 and 06:00, and intervals with more than 20 consecutive minutes without accelerations. Days with more than 480 minutes of active recording were considered valid. School-time PA was defined as occurring between 08:00 and 14:00. Schooldays with more than 40% of active recording during school-time were considered valid. Valid school-time measurements were also determined by manually coding each class’ schedules to control that valid measurements occurred on days when the intervention was scheduled. The subsequent analyses included only participants with at least two

days/schooldays with valid measurements. We used Actilife software (Actigraph, LLC, Pensacola, Florida, USA) to initialize and download the accelerometer data. STATA (Stata Statistical Software, StataCorp LP) was used to process and analyze the raw data. Epoch was set to 10 seconds. Overall PA is expressed as average counts per minute (counts·min^{-1}). To assess average minutes per day and per school-day spent sedentary or in moderate to vigorous PA, time registered with <100 counts per minute and >1999 counts per minute, was divided by valid days/school-days of assessment. We used established cut-points for moderate to vigorous PA of 2000 counts per minute, which is equivalent to a >4km/h walking speed, among adolescents (Kolle, Steene-Johannessen, Andersen & Anderssen, 2010). The interventions' effects on PA are outlined in a separate paper (Kolle et al., 2020, submitted).

Adherence to protocol

Several measures were taken to ensure adherence to protocol. To assess fidelity, adaptation, quality and responsiveness, qualitative process evaluations were carried out on both interventions. The process evaluation for M2 has been published elsewhere (Åvitsland, Ohna, et al., 2020), while the process evaluation for M1, is provisional and only available in Norwegian (Kolle et al., 2019). Dose delivered was also measured: One teacher liaison from each school was responsible for reporting the intervention components as executed/not executed on an online platform. Dose delivered is expressed as the mean percentage of intervention components that were executed relative to the total number of intervention components that were possible to execute during the 29-week intervention period.

2.4 Statistical analyses

All analyses were performed in IBM SPSS Statistics 25 (IBM, Armonk, New York, USA). SDQ data were managed and organized into the predetermined scales by the syntax provided by the SDQ information web page (Youthinmind, 2018). We report descriptive statistics from T1 and T2 as means and standard deviations (SD). We used Cronbach's alpha to assess the internal consistency of TDS and its subscales. The respective results from T1 and T2 were as follows: emotional problems (0.67 and 0.71), conduct problems (0.51 and 0.53), hyperactivity (0.66 and 0.68), peer problems (0.61 and 0.61) and TDS (0.62 and 0.61). We tested the baseline differences between M1 and control, and between M2 and control, using one-way ANOVA with Fisher's LSD post hoc test.

Of the students who consented to participate in the study (n = 2084), 83% (n = 1728) completed the SDQ at T1. Missing values between enrollment and T1 have been described previously (Åvitsland, Leibinger, et al., 2020). Of the completers at T1, 20% (n = 337) did not complete the SDQ at T2, resulting in the complete case group (completers at T1 and T2) including 1391 participants. To examine if the data were missing completely at random (MCAR), Little's MCAR test was conducted with the variables TDS at T1 and T2, sex, SES and immigrant status. The test did not indicate that data were compatible with MCAR (chi square = 39.408; DF = 9; $p < .001$). We used logistic regression to examine whether TDS at T1 or T2, sex, SES, immigrant status and three auxiliary physical fitness variables (cardiorespiratory fitness, muscular strength and body composition; see Åvitsland, Leibinger, et al., 2020), could predict the likelihood of being a complete case or having at least one missing TDS value. The results (odds ratio; 95% CI; p) indicated that missingness did not depend on the outcome (1.0; 0.91 to 1.1; $p = .658$), baseline TDS (1.0; 0.95 to 1.05; $p = .821$) or any other variable included in the model. We therefore assumed the possibility that complete case analyses could produce unbiased results (Hughes, Heron, Sterne, & Tilling, 2019). Two final post hoc tests were carried out to assess whether experimental groups influenced missingness: First, the logistic regression described above was stratified for experimental group to examine whether variables predicted missingness differently between M1, M2 and control. Second, ANOVA with Fisher's LSD post hoc test was used to assess whether TDS at T1, among those with missing TDS at T2, was different between M1, M2 and control. Neither of these tests indicated differences between the groups.

Intervention effects

To test the effect of the interventions, we conducted complete case analyses using linear mixed effects models, with schools as a random effect. We tested the effect of group (M1, M2 and control) on change in the dependent variables, in models that included the respective baseline variable as a covariate. Moderating effects for change in the dependent variables were determined by testing a categorical subgroup * group interaction in models controlling for main effects on group and subgroup. The moderating variables were sex, SES, immigrant status and TDS level at baseline. In cases where the results indicated an interaction effect, subgroups were analyzed separately by stratifying the data file. If a model indicated results that we interpreted to be compatible with an effect on TDS at T2, subsequent analyses were conducted with the TDS subscales as dependent variables. We report the estimated mean difference in change between groups (b), 95% confidence intervals (CI), exact p-values and the intraclass correlation

coefficient (ICC) for the cluster effect of schools. The estimated mean difference in change is expressed by measurement units on the scale of the dependent variable, adjusted for potential baseline differences (M1-control and M2-control).

We have not used the Bonferroni adjustment, as we concur with its critics (Moran, 2003; Nakagawa, 2004; Perneger, 1998) who argued that the adjustment increases the risk of type 2 error, and that "simply describing what tests of significance have been performed, and why, is generally the best way of dealing with multiple comparisons" (Perneger, 1998).

In an effort to adhere to the American Statistical Association's statement on statistical significance and p-values (Wasserstein & Lazar, 2016), and more recent recommendations in a special issue of the journal The American Statistician (Wasserstein, Schirm, & Lazar, 2019); we do not dichotomously interpret the p-values to be either significant or non-significant. Instead, we interpret the p-values as continuous quantities that express how compatible the observed data are with the null-hypotheses: Smaller p-values indicate greater incompatibility with the null-hypotheses. Based on the continuous p-values, the size of the unstandardized regression estimates (b) and the limits of the confidence intervals, we interpret how compatible the results are with our hypotheses and present the results accordingly (Amrhein, Greenland, & McShane, 2019; Greenland et al., 2016). Other factors that influence our interpretations are related to prior evidence, plausibility of mechanism, study design and data quality (McShane, Gal, Gelman, Robert, & Tackett, 2019).

Multiple imputation

Despite the plausible MCAR assumption, the high amount of missing values increases the possibility that the data might be missing at random (MAR) or not missing at random (MNAR). Therefore, multiple imputation was employed. With five imputations and ten iterations, missing data were imputed with TDS at T1 and T2, moderating variables and other auxiliary variables (muscular strength, body composition and cardiorespiratory fitness), using the automatic procedure. We did not impute on the SDQ subscale variables, because SPSS cannot carry out multiple imputation with that many missing values (Mustillo & Kwon, 2015). Analyses that showed results compatible with effects from the complete case data were repeated on the imputed dataset. Complete case results and multiple imputation results are presented as recommended by Manly and Wells (2015) and Sterne et al. (2009).

3. Results

3.1 Descriptives and baseline differences

Table 1 shows the number of participants with valid data from each group at both time points, baseline characteristics and PA levels, results from T1 and T2, and the distribution of participants within the TDS subgroups “normal”, “borderline” or “abnormal” at T1 and T2. The results were compatible with baseline differences between intervention groups and control group: M1 had 6% higher TDS (b = 0.6; 95% CI = 0.003 to 1.17; p = .049) than control, and M2 had 3.4% lower SES (-0.2 to 0.0; p = .039) than control. Furthermore, regarding school-time PA, M1 had 13% fewer counts per minute (b = -65.0; 95% CI = -90.2 to -39.8; p < .001), 6% more sedentary time (b = 14.2; 95% CI = 10.2 to 18.2; p < .001) and 14% less moderate to vigorous PA (b = -4.3; 95% CI = -5.9 to -2.6; p < .001) than control.

Table 1. Participants' demographic characteristics, school-time physical activity at T1 and SDQ scores at T1 and T2. Presented as means with standard deviations (SD)

	M1				**M2**				**Control**			
Demographics	n	%	n		n	%	n		n	%	n	
Sex (% girls/boys)	655	50/50			585	49/51			796	49/51		
Parents' education level (%)	647				583				790			
Lower secondary school or less		6				7				5		
Upper secondary school		27				32				28		
University < 4 years		43				39				42		
University > 4 years		24				22				25		
Non-immigrant (% no/yes)	587	10/90			490	8/92			680	10/90		
Time points		**T1**		**T2**		**T1**		**T2**		**T1**		**T2**
Physical activity (school-time)	484				372				522			
CPM		446.6				511.9				511.5		
Sedentary time (minutes)		233.4				220.3				219.2		
MVPA (minutes)		25.7				29.3				29.9		
Mental health												
Emotional problems (0-10)	592	2.9 (2.1)	577	3.2 (2.3)	484	3.2 (2.2)	453	3.3 (2.4)	654	2.9 (2.2)	556	3.3 (2.5)
Conduct problems (0-10)	592	1.8 (1.5)	577	1.7 (1.4)	484	1.5 (1.5)	453	1.8 (1.6)	654	1.6 (1.6)	556	1.8 (1.7)
Hyperactivity (0-10)	592	4.1 (2.1)	577	4.3 (2.2)	484	4.1 (2.1)	453	4.3 (2.3)	654	3.9 (2.1)	556	4.2 (2.2)
Peer problems (0-10)	592	2.0 (1.7)	577	1.9 (1.7)	484	1.8 (1.6)	453	2.0 (1.8)	654	1.8 (1.7)	555	1.7 (1.7)
TDS (0-40)	590	10.8 (5.2)	577	11.1 (5.2)	484	10.6 (4.9)	453	11.3 (5.6)	654	10.2 (5.4)	555	11.0 (5.7)
TDS cut-offs	n	%	n	%	n	%	n	%	n	%	n	%
Normal	502	85	464	80	400	83	347	77	550	84	451	81
Borderline	49	8	73	13	57	12	64	14	65	10	58	10
Abnormal	39	7	40	7	27	5	42	9	39	6	46	8

Valid n is presented for each variable. School-time physical activity is reported as average minutes per school day for participants with at least two valid days of wear time. CPM = counts per minute; MVPA = moderate to vigorous physical activity; TDS = total difficulties score.

3.2 Adherence to protocol

Average dose delivered in the M1 schools was registered to be 81%, ranging from 72% to 95%. Average dose delivered in the M2 schools was 80%, ranging from 67% to 93%. This is equivalent to 97 and 96 minutes, respectively, out of 120 possible minutes per week of additional school-based PA. The compliance for reporting was 98.2%.

3.3 Intervention effects

Table 2 shows the estimated mean difference in change between the intervention groups and the control group. The ICC was 0.007 for TDS, which indicates small to no difference between schools (Killip, Mahfoud, & Pearce, 2004). For the overall population, the results were incompatible with an effect of M1 or M2 on TDS. We interpreted the results to be incompatible with interaction effects for sex (p = .150) or SES (p = .951), and compatible with interaction effects for immigrant status (p = .061) and baseline TDS levels (p = .008). The subsequent subgroup analyses showed beneficial results. In the abnormal TDS subgroup, results were compatible with a mean difference in change for TDS in favor of M1, compared to their control group counterparts (b = -2.9; 95% CI = -5.73 to -0.07; p = .045). Relative to the estimated baseline levels within the abnormal TDS subgroup (23 points), M1 reduced TDS by 22%, while the control condition reduced TDS by 9%. Subsequent analyses of the SDQ subscales showed that the result could mainly be attributed to difference in change for conduct problems (b = -0.99; 95% CI = -2.02 to 0.04; p = .058) and hyperactivity (b = -1.13; 95% CI = -2.1 to -0.19; p = .019).

In the immigrant subgroup, results were compatible with a mean difference in change for TDS in favor of both M1 (b = -1.6; 95% CI = -3.53 to 0.27; p = .093) and M2 (b = -2.1; 95% CI = -4.36 to 0.21; p = .075), compared to their control group counterparts. Relative to the estimated baseline levels within the immigrant subgroup (11 points), TDS increased 5% in M1, 0% in M2 and 18% in the control group. Subsequent analyses of the SDQ subscales showed that the result in favor of M1 could mainly be attributed to difference in change for emotional problems (b = -1.1; 95% CI = -1.89 to -0.29; p = .008), and the result in favor of M2 could mainly be attributed to difference in change for emotional problems (b = -1.0; 95% CI = -1.99 to -0.07; p = .036) and hyperactivity (b = -0.94; 95% CI = -1.90 to 0.02; p = .055).

The analyses on the immigrant/non-immigrant subgroups with the SDQ subscales as dependent variables showed an unexpected result: in the non-immigrant subgroup, there was compatibility with a mean difference in change for peer problems indicating an increase from M2 (b = 0.32; 95% CI = 0.03 to 0.62; p = .034), compared to the corresponding control subgroup. This warranted further investigation into the subscale peer problems to understand whether there were heterogeneous effects between specific subgroups. The results were compatible with a mean difference in change indicating that M2 increased peer problems among non-immigrant girls (b = 0.42; 95% CI = 0.11 to 0.72; p = .010) and in the subgroup with borderline TDS at baseline (b = 0.89; 95% CI = 0.1 to 1.67; p = .029), compared to their respective control group counterparts. Relative to the estimated baseline levels within the non-immigrant girls subgroup (1.6 points), peer problems increased 19% in M2, while their control group counterparts displayed a 13% decrease. Similarly, relative to estimated baseline levels within the borderline TDS at T1 subgroup, (3.2 points), peer problems increased 6% for M2, while their control counterparts displayed a 31% decrease.

Table 2
Estimated mean difference in change in dependent variables, between groups.

Dependent variable	Groups/subgroup	Estimated mean difference in change (b (95% CI; p))	
		M1-control	M2-control
Total difficulties score			
	All participants (n = 1391)	0.04 (-0.65 to 0.73; p = .899)	0.16 (-0.57 to 0.90; p = .641)
	Normal TDS at T1 (n = 1182)	0.33 (-0.50 to 1.15; p = .410)	0.24 (-0.64 to 1.11; p = .571)
	Borderline TDS at T1 (n = 138)	-1.08 (-3.03 to 0.88; p = .277)	0.21 (-1.66 to 2.08; p = .825)
	Abnormal TDS at T1 (n = 71)	**-2.9 (-5.73 to -0.07; p = .045)**	-1.4 (-4.87 to 1.99; p = .405
	Immigrant (n = 118)	**-1.6 (-3.53 to 0.27; p = .093)**	**-2.1 (-4.36 to 0.21; p = .075)**
	Non-immigrant (n = 1263)	0.14 (-0.55 to 0.84; p = .665)	0.31 (-0.43 to 1.05; p = .382)
Emotional problems			
	Immigrant (n = 118)	**-1.1 (-1.89 to -0.29; p = .008)**	**-1.0 (-1.99 to -0.07; p = .036)**
	Non-immigrant (n = 1263)	0.11 (-0.26 to 0.47; p = .556)	-0.03 (-0.41 to 0.36; p = .892)
Conduct problems			
	Normal TDS at T1 (n = 1182)	0.02 (-0.17 to 0.21; p = .835)	0.02 (-0.19 to 0.23; p = .812)
	Borderline TDS at T1 (n = 138)	-0.01 (-0.77 to 0.74; p = .973)	-0.04 (-0.79 to 0.72; p = .922)
	Abnormal TDS at T1 (n = 71)	**-0.99 (-2.02 to 0.04; p = .058)**	-0.08 (-1.31 to 1.16; p = .903)
Hyperactivity			
	Normal TDS at T1 (n = 1182)	0.22 (-0.05 to 0.48; p = .098)	-0.02 (-0.30 to 0.26; p = .880)
	Borderline TDS at T1 (n = 138)	-0.48 (-1.26 to 0.29; p = .220)	0.06 (-0.69 to 0.80; p = .879)
	Abnormal TDS at T1 (n = 71)	**-1.13 (-2.1 to -0.19; p = .019)**	**-1.07 (-2.22 to 0.08; p = .067)**
	Immigrant (n = 118)	-0.58 (-1.42 to 0.27; p = .171)	**-0.94 (-1.90 to 0.02; p = .055)**
	Non-immigrant (n = 1263)	0.12 (-0.11 to 0.35; p = .292)	0.01 (-0.24 to 0.25; p = .957)
Peer problems			
	Normal TDS at T1 (n = 1182)	0.11 (-0.16 to 0.38; p = .416)	0.24 (-0.04 to 0.53; p = .093)
	Borderline TDS at T1 (n = 138)	0.25 (-0.53 to 1.03; p = .509)	**0.89 (0.10 to 1.67; p = .029)**
	Abnormal TDS at T1 (n = 71)	-0.07 (-1.14 to 1.0; p = .899)	0.26 (-1.01 to 1.53; p = .685)
	Immigrant girls (n = 63)	0.49 (-0.55 to 1.52; p = .349)	0.11 (-1.07 to 1.28; p = .860)
	Non-immigrant girls (n = 652)	0.23 (-0.06 to 0.52; p = .116)	**0.42 (0.11 to 0.72; p = .010)**
	Immigrant boys (n = 55)	-0.14 (-1.11 to 0.83; p = .775)	-0.01 (-1.26 to 1.25; p = .990)
	Non-immigrant boys (n = 612)	-0.03 (-0.44 to 0.38; p = .875)	0.24 (-0.19 to 0.68; p = .260)

Note. We display results from analyses of intervention groups as a whole, and subgroups that show heterogeneous effects. TDS = total difficulties score. T1 = baseline. M1 = intervention group 1. M2 = intervention group 2. Results interpreted to be compatible with effects are accentuated in **bold**.

3.4 Multiple imputation

Compared to the complete case results, the linear mixed effects model conducted on the imputed dataset showed results that were less compatible with effects on TDS, although the unstandardized coefficients and confidence intervals showed similar tendencies (Table 3). These results may be biased, however, because of the large amount of data that were imputed and the majority of missing data existed in the outcome variable (Hughes et al., 2019; Lee & Carlin, 2012). Additionally, multiple imputation on datasets containing cluster randomized groups and subgroups may skew the imputed values toward the mean (Sullivan, White, Salter, Ryan, & Lee, 2018). This may explain why the results for the immigrant subgroup and the

abnormal TDS subgroup were less compatible with effects than in the complete case analyses. For these reasons, the emphasis in the discussion will be placed on the complete case results.

Table 3. Estimated mean difference in change in Total Difficulties Score between groups, using the imputed dataset

Dependent variable	Groups analyzed	Estimated mean difference in change (b (95% CI; p)	
		M1-control	M2-control
TDS			
	All participants (n = 2045)	-0.04 (-0.6 to 0.5; p = .884)	0.08 (-0.6 to 0.7; p = .793)
	Normal TDS at T1 (n = 1452)	0.26 (-0.4 to 1.0; p = .461)	0.20 (-0.5 to 0.9; p = .569)
	Borderline TDS at T1 (n = 171)	-1.22 (-3.0 to 0.5; p = .176)	0.34 (-1.3 to 2.0; p = .693)
	Abnormal TDS at T1 (n = 105)	-1.68 (-4.2 to 0.83; p = .188)	-0.98 (-4.0 to 2.0; p = .519)
	Immigrants (n = 192)	-0.94 (-2.9 to 1.0; p = .340)	-1.1 (-3.2 to 0.9; p = .280)
	Non-immigrants (n = 1893)	0.07 (-0.4 to 0.6; p = .775)	0.32 (-0.2 to 0.9; p = p .252)

Note. T1 = baseline. M1 = intervention model 1. M2 = intervention model 2.

4. Discussion

The aim of this paper was to assess the effect of two school-based PA interventions on adolescents' mental health. The complete case results indicate that the interventions did not affect TDS in the overall population. Subgroup analyses, however, showed beneficial effects from both interventions. Specifically, M1 reduced TDS in the subgroup with the highest levels of psychological difficulties at T1 and both interventions prevented an increase in TDS for a majority in the immigrant subgroup. Analyses of the SDQ subscales revealed, surprisingly, that M2 caused peer problems to increase in both the non-immigrant girls subgroup and the borderline TDS at T1 subgroup.

Although the immigrant subgroups' mixed model p-values were non-significant in the traditional sense, they were low enough to indicate that the data conformed more to the hypothesis of an effect, than the null hypothesis (Greenland et al., 2016). Furthermore, compared to M1, the difference in change for 95% of the control group spanned from a 3.5 points bigger increase to a 0.3 points lower increase. Compared to M2, the difference in change for 95% of the control group spanned from a 4.4 points bigger increase to a 0.2 points lower increase. The two upper limits indicate an increase with no practical implications, while the lower limits indicate substantial clinically significant effects. This interpretation is based on the findings by Goodman and Goodman (2009), suggesting that every one-point increase in TDS represents a 16% to 23% increased likelihood of developing a mental disorder. Therefore, even though the p-values were above the traditional significance level and the confidence interval contained the null-value; a majority of the immigrant subgroup who received the interventions

may have experienced a substantial decrease in the likelihood of developing a mental disorder, compared to the respective control subgroup.

4.1 Overall population

The lack of an overall effect was not surprising, considering that 85% and 83% of the respective M1 and M2 populations had normal levels of psychological difficulties. Similar ceiling effects were suspected in comparable studies by Eather et al. (2016), Smith et al. (2018) and Christiansen et al. (2018). Although our intervention period of 29 weeks was longer than in similar studies (Eather et al., 2016; Lubans et al., 2016; Smith et al., 2018); it might have been too short to attenuate an overall increase in psychological difficulties among generally healthy adolescents. Furthermore, although 80-81% of the dose was registered as delivered, analyses on PA outcomes show that M1 slightly increased school-time PA levels, while M2 did not (Kolle et al., 2020). This difference may have been caused by the design of the M2 intervention: All three M1 intervention components were teacher-led and were anticipated to be performed with moderate to high intensity. M2, however, contained two lessons that were mainly student-led. The process evaluation substantiates the hypothesis that the extensive freedom and student-led activities that characterized M2 sometimes led to truancy (Åvitsland, Ohna, et al., 2020). Additionally, the intervention specified that students were allowed to choose their preferred activity, which involved everything from low intensity walking to high intensity soccer. These factors may have resulted in a dichotomization among the M2 participants, characterized by physical activity and inactivity. It is also important to note that the control population was not physically inactive: The Norwegian lower secondary school curriculum mandates at least two PE lessons per week and students can also opt in for the elective subject of physical activity and health (often organized in weekly 90-minute lessons). In addition, 63% of lower secondary school students in Norway participate in leisure time sports (Bakken, 2019). Among upper secondary school students, however, only 40% participate in leisure time sports and it is possible that an intervention continuing into this period of adolescence would have shown an effect in the overall population.

4.2 Subgroups

Abnormal TDS at T1

TDS decreased in all abnormal TDS at T1-subgroups respective to M1, M2 and control, indicating a regression to the mean. However, the abnormal TDS at T1 subgroup that received the M1 intervention displayed a reduction in TDS that was more than twice as big as the reduction in the corresponding control subgroup. The almost 3 points larger mean reduction in M1 may be a substantial clinically significant difference (Goodman & Goodman, 2009). The results concur with previous studies that also found effects in similar subgroups (Christiansen et al., 2018; Eather et al., 2016; Smith et al., 2018).

The effect on TDS could be attributed to reductions in conduct problems and hyperactivity, although this does not align with the meta-analysis by Ahn and Fedewa (2011), which did not show an association between PA and conduct problems. However, a possible explanation for the present reduction is that PA can influence mental health through a behavioral mechanism, for example by improving coping and self-regulation (Lubans et al., 2016), which are inversely associated with conduct problems in adolescents (Ebata & Moos, 1991). The reduction in hyperactivity is supported by a substantial amount of evidence suggesting that PA has a beneficial effect on hyperactivity through neurobiological pathways (Gapin, Labban, & Etnier, 2011).

Immigrants

The immigrant subgroups that received the M1 and M2 interventions displayed a smaller increase and no increase in TDS, respectively, compared to the corresponding control subgroup. Mainly, the effect from M1 could be attributed to reductions in emotional problems, and the effect from M2 could be attributed to reductions in emotional problems and a prevented increase in hyperactivity. These effects might be explained by a psychosocial mechanism (Lubans et al., 2016): Social support may be the psychosocial mechanism that was affected, as it has been shown to be associated with sports participation (Babiss & Gangwisch, 2009) and the M2 intervention was designed to facilitate positive social relationships. Immigrant adolescents experience less social support than non-immigrants, and therefore might have a larger potential for change (Oppedal & Røysamb, 2004). In turn, social support can stave off emotional problems (Garnefski & Diekstra, 1996) and is inversely associated with hyperactivity

(Mastoras, Saklofske, Schwean, & Climie, 2015). Self-esteem is another potential psychosocial mechanism that may explain the results, considering that PA is associated with adolescent self-esteem (Dale, Vanderloo, Moore, & Faulkner, 2019), which, in turn is associated with hyperactivity (Edbom, Lichtenstein, Granlund, & Larson, 2006) and emotional health (Moksnes & Espnes, 2012). Moreover, adolescent immigrants may have lower self-esteem than their non-immigrant peers (Bankston & Zhou, 2002). To discuss potential reasons why immigrant adolescents experience less social support or have poorer senses of selves than their non-immigrant peers is beyond the scope of this paper. However, racial discrimination occurs, and can influence connected outcomes such as perceived physical appearance, feelings of belonging to a peer group and identity development (Virta, Sam, & Westin, 2004). To the authors' knowledge, no previous study on the effect of school-based PA interventions on mental health in adolescent populations has specifically identified immigrants as a subgroup. However, in a recent study on an adult immigrant population in Sweden, a four-month lifestyle intervention positively influenced mental health, and the authors emphasized increased PA and social support as potential causes (Siddiqui, Lindblad, Nilsson, & Bennet, 2019). Furthermore, in a qualitative study, "enhanced self-confidence, happiness, and lower stress" were frequently reported as experienced benefits from PA among adult and adolescent immigrants of different origins in USA (Wieland et al., 2015). The immigrant experience of school-based PA interventions has been neglected in previous research. Similar studies in the future should identify this subgroup, not only to assess quantitative effects, but also to assess why this subgroup may benefit from increasing school-time PA.

Non-immigrant girls and borderline TDS at T1

In the subgroups non-immigrant girls and borderline TDS at T1 that received the M2 intervention, peer problems increased, while their respective control counterparts displayed decreases. Although the differences in change were small to moderate – respectively 0.4 and 0.9 points – the percentage results relative to baseline gave reason for concern. Scoring high on peer problems is characterized by the SDQ as being solitary, having few friends, being bullied by others and getting along better with adults than with peers. The negative effects were surprising, considering that the intervention was designed to facilitate social relationships through PA. Additionally, participation in team sports has been associated with fewer mental problems than participation in individual sports (Breistøl, Clench-Aas, Van Roy, & Raanaas, 2017), and a recent review by Pels and Kleinert (2016) concluded that PA could contribute to reducing loneliness. The process evaluation of the M2 intervention offers a possible explanation

for the negative effect among girls (Åvitsland, Ohna, et al., 2020): The formation of activity groups could lead to some girls feeling ostracized, while planning and cooperating within the group sometimes led to disagreements, which could cause one or several girls to leave their groups. Previous research on sex differences in adolescent peer relationships has shown that social anxiety, expressed by a "fear of negative evaluation from peers, and more social avoidance and distress in new situations" (La Greca & Lopez, 1998), is more prevalent in girls than in boys. Furthermore, while boys tend to thrive in social groups, girls tend to place more emphasis on dyadic relationships (Prinstein, Borelli, Cheah, Simon, & Aikins, 2005), perhaps because they experience less conflicts in these relationships. According to Xie, Swift, Cairns, and Cairns (2002), conflicts among girls involving social aggression, e. g., exclusion, isolation and gossiping, most often occur in groups of four or more members, a common group size in the M2 intervention. The potential sources of conflict for the non-immigrant girls may also have been the reason for the negative effect in the borderline TDS subgroup; however, it is unclear why the negative effect occurred specifically in this subgroup.

4.3 Strengths and limitations

Strengths of this study include a large sample of an understudied population, the use of a cluster-RCT design and multilevel analyses of whole groups and subgroups. The mental health outcome variable is comprised of four subscales, which can be helpful for interpreting the explanatory mechanisms. Except for the non-immigrant girls subgroup, all subgroups were determined a priori and fulfil most of the credibility criteria set by Sun et al. (2012). However, limitations must be addressed and there are at least six specific factors that increase the uncertainty of the results: 1) The participating schools volunteered and may be systematically different from schools that declined to participate. This may restrict the generalizability of the results. 2) There is also the possibility of a collider bias, i. e. that the mental health of participants predicted different rates of participation at follow-up, depending on the experimental groups. Between T1 and T2, M1, M2 and the control group lost 22%, 36% and 40% of their respective participants. The loss in M2 can be explained by the school withdrawing from the project. The high attrition in the control group, however, cannot be accounted for and may influence the results somehow. 3) Although we assumed that complete case analyses could produce unbiased results, the assumption is uncertain. The high levels of missing values may have influenced the results. 4) The subgroups in which we interpreted effects contained between 5% and 47% of the full population and testing of the smallest subgroups may have been low in power. The small subgroups may also not be representative of equivalent subgroups outside of

the study population. 5) The internal consistencies of TDS and subscales were below the recommended cutoff point at 0.7 (Bland & Altman, 1997). This could be due to poor understanding of the questions, unwillingness to answer the questions honestly, or actual low consistency. Although this contributes to uncertainty, it should be noted that the relevance of Cronbach's alpha has been criticized (Sijtsma, 2009). It should also be considered a limitation that we only used one measure of mental health, as there may be other instruments that are more sensitive to change. 6) Lastly, schools are complex contexts (Moore et al., 2019) and our interventions can be characterized as complex (Craig et al., 2008). This means that there are many potential interacting systems that we cannot control, that may influence the results.

5. Conclusions

The School in Motion cluster-RCT with two intervention arms spanning over 29 weeks, did not affect psychological difficulties in the overall population (14-15-year-olds in Norway attending ninth grade). Results indicated beneficial effects in two subgroups: those with the highest baseline levels of psychological difficulties and immigrants. The effects could be attributed either to a reduction, or prevented increase, in conduct problems, hyperactivity and/or emotional problems. The present results indicate that school-based PA interventions may cause clinically significant changes in psychological difficulties in these subgroups, and these changes may reduce the odds of developing mental disorders. The M2 intervention may be beneficial for immigrants in its current form, although further studies are needed with adapted versions to avoid an increase in peer problems in some subgroups. Future research should focus on the causal relationship between school-time PA and adolescent immigrants' mental health, as no previous research exists on this subject.

References

Abebe, D. S., Lien, L., & Hjelde, K. H. (2014). What We Know and Don't Know About Mental Health Problems Among Immigrants in Norway. *Journal of Immigrant Minority Health, 16*(1), 60-67. https://doi.org/10.1007/s10903-012-9745-9

Agans, J. P., Sävfenbom, R., Davis, J. L., Bowers, E. P. & Lerner, R. M. (2013). Positive Movement Experiences: Approaching the Study of Athletic Participation, Exercise, and Leisure Activity through Relational Developmental Systems Theroy and the Concept of Embodiment. *Advances in Child Development and Behavior, 45,* 261-286. 10.1016/B978-0-12-397946-9.00010-5

Ahn, S., & Fedewa, A. L. (2011). A Meta-analysis of the Relationship Between Children's Physical Activity and Mental Health. *Journal of Pediatric Psychology, 36,* 385-397. doi:10.1093/jpepsy/jsq107

Amrhein, V., Greenland, S., & McShane, B. (2019). Scientists rise up against statistical significance. *Nature, 567,* 305-307. 10.1038/d41586-019-00857-9

Babiss, L. A., & Gangswisch, J. E. (2009) Sports participation as a protective factor against depression and suicidal ideation in adolescents as mediated by self-esteem and social support. *Journal of Developmental & Behavioral Pediatrics 30*(5), 376-384. 10.1097/DBP.0b013e3181b33659

Bailey, A. P., Hetrick, S. E., Rosenbaum, S., Purcell, R., & Parker, A. G. (2017). Treating depression with physical activity in adolescents and young adults: a systematic review and meta-analysis of randomised controlled trials. *Psychol Med, 48*(7), 1-20. 10.1017/S0033291717002653

Bakken, A (2015) *Ungdata: Nasjonale resultater 2014.* (Nova report no. 7/15) Oslo: Høgskolen i Oslo og Akershus. Retrieved from http://www.hioa.no/Om-OsloMet/Senter-for-velferds-og-arbeidslivsforskning/NOVA/Publikasjonar/Rapporter/2015/Ungdata.-Nasjonale-resultater-2014

Bakken, A (2019). *Ungdata. Nasjonale resultater 2019.* Nova report no 9/19) Oslo: Oslo MET. Retrieved from https://fagarkivet.oslomet.no/en/item/ungdata-2019-nasjonale-resultater

Bandura. A. (1982). Self-efficacy mechanism in human agency. *American Psychologist, 37*(2), 122- 147. 10.1037/0003-066X.37.2.122

Bankston, C. L., III & Zhou, M. (2002). Being well vs. doing well: Self-esteem and school performance among immigrant and nonimmigrant racial and ethnic groups. *The International Migration Review, 36*(2), 389-415. http://dx.doi.org/10.1111/j.1747-7379.2002.tb00086.x

Biddle, S. J. H., Ciaccioni, S., Thomas, G., & Vergeer, I. (2019). Physical activity and mental health in children and adolescents: An updated review of reviews and an analysis of causality. *Psychology of Sport and Exercise, 42,* 146-155. https://doi.org/10.1016/j.psychsport.2018.08.011

Bland, J. M., & Altman, D. G. (1997). Statistics notes: Cronbach's alpha. *BMJ, 314*(7080), 572. 10.1136/bmj.314.7080.572

Bonhauser, M., Fernandez, G., Püschel, K., Yañez, F., Montero, J., Thompson, B., & Coronado, G. (2005). Improving physical fitness and emotional well-being in

adolescents of low socioeconomic status in Chile: results of a school-based controlled trial. *Health Promotion International, 20*(2), 113-122. 10.1093/heapro/dah603

Breistøl, S., Clench-Aas, J., Van Roy, B., & Raanaas, R. K. (2017). Association Between Participating in Noncompetitive or Competitive Sports and Mental Health among Adolescents – a Norwegian Population- based Cross- sectional Study. *Scandinavian Journal of Child and Adolescent Psychiatry and Psychology, 5*(1), 28-38. 10.21307/sjcapp-2017-003

Bøe, T., Øverland, S., Lundervold, A. J., & Hysing, M. (2012). Socioeconomic status and children's mental health: results from the Bergen Child Study. *Soc Psychiatry Psychiatr Epidemiol, 47*(10), 1557-1566. https://doi.org/10.1007/s00127-011-0462-9

Casey, M. M., Harvey, J. T., Telford, A., Eime, R. M., Mooney, A., & Payne, W. R. (2014). Effectiveness of a school- community linked program on physical activity levels and health- related quality of life for adolescent girls. *BMC Public Health, 14*, 649-649. 10.1186/1471-2458-14-649

Cerin, E. (2010). Ways of unraveling how and why physical activity influences mental health through statistical mediation analyses. *Mental Health and Physical Activity, 3*(2), 51-60. https://doi.org/10.1016/j.mhpa.2010.06.002

Christiansen, L. B., Lund-Cramer, P., Brondeel, R., Smedegaard, S., Holt, A.-D., & Skovgaard, T. (2018). Improving children's physical self-perception through a school-based physical activity intervention: The Move for Well-being in School study. *Mental Health and Physical Activity, 14*, 31-38. 10.1016/j.mhpa.2017.12.005

Collishaw, S. (2015). Annual Research Review: Secular trends in child and adolescent mental health. *Journal of Child Psychology and Psychiatry, 56*(3), 370-393. 10.1111/jcpp.12372

Craig, P., Dieppe, P., Macintyre, S., Michie, S., Nazareth, I., & Petticrew, M. (2008). Developing and evaluating complex interventions: the new Medical Research Council guidance. *BMJ, 337*, a1655. 10.1136/bmj.a1655

Dale, L. P., Vanderloo, L., Moore, S. & Faulkner, S. (2019). Physical activity and depression, anxiety and self-esteem in children and youth: An umbrella systematic review. *Mental Health and Physical Activity, 16*, 66-79. https://doi.org/10.1016/j.mhpa.2018.12.001

Dalene, K. E., Anderssen, S. A., Andersen, L. B., Steene-Johannessen, J., Ekelund, U., Hansen, B. H., & Kolle, E. (2018). Secular and longitudinal physical activity changes in population-based samples of children and adolescents. *Scandinavian journal of medicine & science in sports, 28*(1), 161-171. 10.1111/sms.12876

Donnelly, E. J., Hillman, H. C., Castelli, L. D., Etnier, N. J., Lee, N. S., Tomporowski, N. P., . . . Szabo-Reed, N. A. (2016). Physical Activity, Fitness, Cognitive Function, and Academic Achievement in Children: A Systematic Review. *Medicine & Science in Sports & Exercise, 48*(6), 1223-1224. 10.1249/MSS.0000000000000966

Eather, N., Morgan, P. J., & Lubans, D. R. (2016). Effects of exercise on mental health outcomes in adolescents: Findings from the CrossFit™ teens randomized controlled trial. *Psychology of Sport and Exercise, 26*, 14-23. https://doi.org/10.1016/j.psychsport.2016.05.008

Ebata, A. T., & Moos, R. H. (1991). Coping and adjustment in distressed and healthy adolescents. *Journal of Applied Developmental Psychology, 12*(1), 33-54. https://doi.org/10.1016/0193-3973(91)90029-4

Erola, J., Jalonen, S., & Lehti, H. (2016). Parental education, class and income over early life course and children's achievement. *Research in Social Stratification and Mobility, 44*, 33-43. https://doi.org/10.1016/j.rssm.2016.01.003

Erskine, H. E., Moffitt, T. E., Copeland, W. E., Costello, E. J., Ferrari, A. J., Patton, G., . . . Scott, J. G. (2015). A heavy burden on young minds: the global burden of mental and substance use disorders in children and youth. *Psychol Med, 45*(7), 1551-1563. http://dx.doi.org/10.1017/s0033291714002888

Fröjd, S. A., Nissinen, E. S., Pelkonen, M. U. I., Marttunen, M. J., Koivisto, A.-M., & Kaltiala-Heino, R. (2008). Depression and School Performance in Middle Adolescent Boys and Girls. *Journal of Adolescence, 31*(4), 485-498. 10.1016/j.adolescence.2007.08.006

Gapin, J. I., Labban, J. D., & Etnier, J. L. (2011). The effects of physical activity on attention deficit hyperactivity disorder symptoms: The evidence. *Preventive Medicine, 52*, S70-S74. https://doi.org/10.1016/j.ypmed.2011.01.022

Garnefski, N., & Diekstra, R. F. W. (1996). Perceived Social Support from Family, School, and Peers: Relationship with Emotional and Behavioral Problems among Adolescents. *Journal of the American Academy of Child & Adolescent Psychiatry, 35*(12), 1657-1664. https://doi.org/10.1097/00004583-199612000-00018

Glied, S., & Pine, D. S. (2002). Consequences and Correlates of Adolescent Depression. *JAMA Pediatrics, 156*(10), 1009-1014. https://doi.org/10.1001/archpedi.156.10.1009

Goodman, A., & Goodman, R. (2009). Strengths and Difficulties Questionnaire as a Dimensional Measure of Child Mental Health. *Journal of the American Academy of Child & Adolescent Psychiatry, 48*(4), 400-403. https://doi.org/10.1097/CHI.0b013e3181985068

Goodman, R. (1997). The Strengths and Difficulties Questionnaire: A Research Note. *Journal of Child Psychology and Psychiatry, 38*(5), 581-586. 10.1111/j.1469-7610.1997.tb01545.x

Goodman, R., Ford, T., Simmons, H., Gatward, R., & Meltzer, H. (2000). Using the Strengths and Difficulties Questionnaire (SDQ) to screen for child psychiatric disorders in a community sample. *The British journal of psychiatry : the journal of mental science, 177*, 534. https://doi.org/10.1192/bjp.177.6.534

Greenland, S., Senn, S. J., Rothman, K. J., Carlin, J. B., Poole, C., Goodman, S. N., & Altman, D. G. (2016). Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations. *European Journal of Epidemiology, 31*(4), 337-350. https://doi.org/10.1007/s10654-016-0149-3

Heelan, K., Welk, G., Pfeiffer Karin, A., Eisenmann Joey, C., Drenowatz, C., Gentile, D., & Walsh, D. (2010). Influence of socio-economic status on habitual physical activity and sedentary behavior in 8- to 11-year old children. *BMC Public Health, 10*(1), 214. 10.1186/1471-2458-10-214

Hills, A. P., Dengel, D. R., & Lubans, D. R. (2015). Supporting Public Health Priorities: Recommendations for Physical Education and Physical Activity Promotion in Schools. *Progress in Cardiovascular Diseases, 57*(4), 368-374. https://doi.org/10.1016/j.pcad.2014.09.010

Hughes, R. A., Heron, J., Sterne, J. A. C., & Tilling, K. (2019). Accounting for missing data in statistical analyses: multiple imputation is not always the answer. *International Journal of Epidemiology, 48*(4), 1294-1304. https://doi.org/10.1093/ije/dyz032

Kessler, R. C., Amminger, G. P., Aguilar-Gaxiola, S., Alonso, J., Lee, S., & Ustun, T. B. (2007). Age of onset of mental disorders: A review of recent literature. *Curr Opin Psychiatry, 20*(4), 359-364. http://dx.doi.org/10.1097/YCO.0b013e32816ebc8c

Killip, S., Mahfoud, Z., & Pearce, K. (2004). What is an intracluster correlation coefficient? Crucial concepts for primary care researchers. *The Annals of Family Medicine, 2*(3), 204-208. https://doi.org/10.1370/afm.141

Kinge, J. M., Sælensminde, K., Dieleman, J., Vollset, S. E., & Norheim, O. F. (2017). Economic losses and burden of disease by medical conditions in Norway. *Health Policy, 121*(6), 691-698. https://doi.org/10.1016/j.healthpol.2017.03.020

Kolle, E., Steene-Johannessen, J., Andersen, L. B., & Anderssen, S. A. (2010) Objectively assessed physical activity and aerobic fitness an a population-based sample of Norwegian 9- and 15-year-olds. *Scandinavian Journal of Medicine & Science in Sports, 20*(1), e41-e47. 10.1111/j.1600-0838.2009.00892.x

Kolle, E., Steene-Johannessen, J., Sävfenbom, R., Grydeland, M., Tjomsland, H. E., Berntsen, S., . . . Solberg, R. B. (2019) *School in Motion – Main Report* (Report no. unknown) Oslo: The Norwegian Directorate for Education and Training. Retrieved from https://www.udir.no/contentassets/00554e6be9104daeb387287132cef1e0/sluttrapport-scim.pdf

Kolle, E., Solberg, R., Säfvenbom, R., Berntsen, S., Resaland, G. K., Ekelund, U., . . . Grydeland, M. (2020) Change in physical activity and aerobic fitness after a school-based intervention among adolescents: The school in motion cluster randomized trial. *Manuscript submitted for publication*

La Greca, A. M., & Lopez, N. (1998). Social Anxiety Among Adolescents: Linkages with Peer Relations and Friendships. *Journal of Abnormal Child Psychology, 26*(2), 83-94. https://doi.org/10.1023/A:1022684520514

Lee, K. J., & Carlin, J. B. (2012). Recovery of information from multiple imputation: a simulation study. *Emerging Themes in Epidemiology, 9*(1), 3. https://doi.org/10.1186/1742-7622-9-3

Lerner, R. M., Hershberg, R., Hillard, L. & Johnson, S. (2015). Concepts and Theories in Human Development. In M. H. Bornstein & M. E. Lamb (Eds.) *Developmental science: An advanced textbook* (pp. 3-42) New York: Psychology Press.

Lerner, R. M. & Lerner, J. V. (2013) *The Positive Development of Youth: Comprehensive findings from the 4-H study of positive youth development.* Retrieved from Washington, Maryland: https://4-h.org/wp-content/uploads/2016/02/4-H-Study-of-Positive-Youth-Development-Full-Report.pdf

Lubans, D. R., Richards, J., Hillman, C., Faulkner, G., Beauchamp, M., Nilsson, M., . . . Biddle, S. (2016). Physical Activity for Cognitive and Mental Health in Youth: A Systematic Review of Mechanisms. *Pediatrics, 138*(3), e20161642. 10.1542/peds.2016-1642

Lubans, D. R., Smith, J. J., Morgan, P. J., Beauchamp, M. R., Miller, A., Lonsdale, C., . . . Dally, K. (2016). Mediators of Psychological Well-being in Adolescent Boys. *Journal of Adolescent Health, 58*(2), 230-236. https://doi.org/10.1016/j.jadohealth.2015.10.010

Manly, C. A., & Wells, R. S. (2015). Reporting the Use of Multiple Imputation for Missing Data in Higher Education Research. *Research in Higher Education, 56*(4), 397-409. https://doi.org/10.1007/s11162-014-9344-9

Mastoras, S., Saklofske, D., Schwean, V., & Climie, E. (2015). Social Support in Children With ADHD: An Exploration of Resilience. *Journal of attention disorders, 22*. 10.1177/1087054715611491

McShane, B. B., Gal, D., Gelman, A., Robert, C., & Tackett, J. L. (2019). Abandon Statistical Significance. *The American Statistician, 73*(sup1), 235-245. https://doi.org/10.1080/00031305.2018.1527253

Moksnes, U. K., & Espnes, G. A. (2012) Self-esteem and emotional health in adolescents – gender and age as potential moderators. *Scandinavian Journal of Psychology, 53*(6), 483-489. 10.1111/sjop.12021

Moore, G. F., Evans, R. E., Hawkins, J., Littlecott, H., Melendez-Torres, G. J., Bonell, C., & Murphy, S. (2019). From complex social interventions to interventions in complex social systems: Future directions and unresolved questions for intervention development and evaluation. *Evaluation, 25*(1), 23-45. 10.1177/1356389018803219

Moran, M. D. (2003). Arguments for rejecting the sequential Bonferroni in ecological studies. *100*(2), 403-405. 10.1034/j.1600-0706.2003.12010.x

Muris, P. (2002). Relationships between self-efficacy and symptoms of anxiety disorders and depression in a normal adolescent sample. *Personality and Individual Differences, 32*(2), 337-348. https://doi.org/10.1016/S0191-8869(01)00027-7

Mustillo, S., & Kwon, S. (2015). Auxiliary Variables in Multiple Imputation When Data Are Missing Not at Random. *The Journal of Mathematical Sociology, 39*(2), 73-91. https://doi.org/10.1080/0022250X.2013.877898

Nakagawa, S. (2004). A farewell to Bonferroni: the problems of low statistical power and publication bias. *Behavioral Ecology, 15*(6), 1044-1045. https://doi.org/10.1093/beheco/arh107

Oppedal, B., & Røysamb, E. (2004). Mental health, life stress and social support among young Norwegian adolescents with immigrant and host national background. *Scandinavian Journal of Psychology, 45*(2), 131-144. 10.1111/j.1467-9450.2004.00388.x

Pels, F., & Kleinert, J. (2016). Loneliness and physical activity: A systematic review. *International Review of Sport and Exercise Psychology, 9*, 1-30. 10.1080/1750984X.2016.1177849

Perneger, T. V. (1998). What's wrong with Bonferroni adjustments. *BMJ, 316*(7139), 1236-1238. 10.1136/bmj.316.7139.1236

Pinto Pereira, S. M., Geoffroy, M. & Power, C. (2014). Depressive symptoms and physical activity during 3 decades in adult life: Bidirectional associations in a prospective cohort study. *JAMA Psychiatry, 71*(12), 1373-1380. 10.1001/jamapsychiatry.2014.1240

Prinstein, M. J., Borelli, J. L., Cheah, C. S. L., Simon, V. A., & Aikins, J. W. (2005). Adolescent Girls' Interpersonal Vulnerability to Depressive Symptoms: A Longitudinal Examination of Reassurance-Seeking and Peer Relationships. *Journal of Abnormal Psychology, 114*(4), 676-688. 10.1037/0021-843X.114.4.676

Ryan, R. M. & Deci, E. L. (2002). An overview of Self-determination Theory: An organismic-dialectical perspective. In E. L. D. R. M. Ryan (Ed.), *Handbook of Self-determination Research* (pp. 3-33). Rochester, New York: The University of Rochester Press.

Sagatun, Å., Kolle, E., Anderssen, S. A., Thoresen, M., & Søgaard, A. J. (2008). Three-year follow-up of physical activity in Norwegian youth from two ethnic groups: associations with socio-demographic factors. *BMC Public Health, 8*(1), 419. https://doi.org/10.1186/1471-2458-8-419

Siddiqui, F., Lindblad, U., & Bennet, L. (2014). Physical inactivity is strongly associated with anxiety and depression in Iraqi immigrants to Sweden: a cross-sectional study. *BMC Public Health, 14*(1), 502. https://doi.org/10.1186/1471-2458-14-502

Siddiqui, F., Lindblad, U., Nilsson, P. M., & Bennet, L. (2019). Effects of a randomized, culturally adapted, lifestyle intervention on mental health among Middle-Eastern immigrants. *European Journal of Public Health, 29*(5), 888-894. https://doi.org/10.1093/eurpub/ckz020

Sijtsma, K. (2009). On the Use, the Misuse, and the Very Limited Usefulness of Cronbach's Alpha. *Psychometrika, 74*(1), 107-120. 10.1007/s11336-008-9101-0

Singh, A. S., Saliasi, E., van den Berg, V., Uijtdewilligen, L., de Groot, R. H. M., Jolles, J., . . . 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), 640-647. 10.1136/bjsports-2017-098136

Singh, G. K., Yu, S. M., Siahpush, M., & Kogan, M. D. (2008). High Levels of Physical Inactivity and Sedentary Behaviors Among US Immigrant Children and Adolescents. *Archives of Pediatrics & Adolescent Medicine, 162*(8), 756-763. https://doi.org/10.1001/archpedi.162.8.756

Smith, J., Beauchamp, M., Faulkner, G., Morgan, P., Kennedy, S., & Lubans, D. R. (2018). Intervention effects and mediators of well-being in a school-based physical activity program for adolescents: The 'Resistance Training for Teens' cluster RCT. *Mental Health and Physical Activity, 15*, 88-94. https://doi.org/10.1016/j.mhpa.2018.08.002

Spruit, A., Assink, M., van Vugt, E., van der Put, C., & Stams, G. J. (2016). The effects of physical activity interventions on psychosocial outcomes in adolescents: A meta-analytic review. *Clinical Psychology Review, 45*, 56-71. https://doi.org/10.1016/j.cpr.2016.03.006

Sterne, J. A. C., White, I. R., Carlin, J. B., Spratt, M., Royston, P., Kenward, M. G., . . . Carpenter, J. R. (2009). Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *The BMJ, 338*, b2393. 10.1136/bmj.b2393

Sullivan, T. R., White, I. R., Salter, A. B., Ryan, P., & Lee, K. J. (2018). Should multiple imputation be the method of choice for handling missing data in randomized trials? *Statistical Methods in Medical Research, 27*(9), 2610-2626. 10.1177/0962280216683570

Sun, X., Briel, M., Busse, J. W., You, J. J., Akl, E. A., Mejza, F., . . . Guyatt, G. H. (2012). Credibility of claims of subgroup effects in randomised controlled trials: systematic review. *BMJ, 344*, e1553. 10.1136/bmj.e1553

Van Roy, B., Veenstra, M., & Clench-Aas, J. (2008). Construct validity of the five-factor Strengths and Difficulties Questionnaire (SDQ) in pre-, early, and late adolescence. *J Child Psychol Psychiatry, 49*(12), 1304-1312. 10.1111/j.1469-7610.2008.01942.x

Virta, E., Sam, D. L., & Westin, C. (2004). Adolescents with Turkish background in Norway and Sweden: A comparative study of their psychological adaptation. *Scandinavian Journal of Psychology, 45*(1), 15-25. 10.1111/j.1467-9450.2004.00374.x

Wang, R., & Ware, J. H. (2013). Detecting Moderator Effects Using Subgroup Analyses. *Prev Sci, 14*(2), 111-120. http://dx.doi.org/10.1007/s11121-011-0221-x

Wasserstein, R. L., & Lazar, N. A. (2016). The ASA Statement on p-Values: Context, Process, and Purpose. *The American Statistician, 70*(2), 129-133. https://doi.org/10.1080/00031305.2016.1154108

Wasserstein, R. L., Schirm, A. L., & Lazar, N. A. (2019). Moving to a World Beyond "p < 0.05". *The American Statistician, 73*(sup1), 1-19. https://doi.org/10.1080/00031305.2019.1583913

Whitehead, M. (2010) *Physical Literacy: Throughout the Lifecourse.* Abingdon, Oxon: Routledge.

WHO. (2010). *Global recommendations on physical activity for health.* Switzerland: World Health Organization. Retrieved from https://apps.who.int/iris/bitstream/handle/10665/44399/9789241599979_eng.pdf;jsessionid=2DD612F706358D095E634489C5C63F90?sequence=1

Wieland, M. L., Tiedje, K., Meiers, S. J., Mohamed, A. A., Formea, C. M., Ridgeway, J. L., . . . Sia, I. G. (2015). Perspectives on Physical Activity Among Immigrants and

Refugees to a Small Urban Community in Minnesota. *Journal of Immigrant and Minority Health, 17*(1), 263-275. https://doi.org/10.1007/s10903-013-9917-2

Xie, H., Swift, D. J., Cairns, B. D., & Cairns, R. B. (2002). Aggressive Behaviors in Social Interaction and Developmental Adaptation: A Narrative Analysis of Interpersonal Conflicts During Early Adolescence. *Social Development, 11*(2), 205-224. 10.1111/1467-9507.00195

Youthinmind. (2018). Information for researchers and professionals about the Strengths & Difficulties Questionnaires. Retrieved from http://www.sdqinfo.com/

Åvitsland, A., Leibinger, E., Solberg, R., Lerum, Ø., Haugen, T., Kolle, E., & Dyrstad, S. M. (2020). The Association Between Physical Fitness and Mental Health in Norwegian Adolescents. *BMC Public Health, 20.* doi: https://doi.org/10.1186/s12889-020-08936-7

Åvitsland, A., Ohna, S. E., Dyrstad, S. M., Tjomsland, H. E., Lerum, Ø., & Leibinger, E. (2020). The process evaluation of a school-based physical activity intervention: Influencing factors and potential consequences of implementation. *Health Education, 120*(2). doi: https://doi.org/10.1108/HE-01-2020-000

Paper 4

The association between a one-year change in physical fitness and mental health status in Norwegian adolescents

Manuscript submitted for publication

Andreas Åvitsland[a], Eva Leibinger[a], Elin Kolle[b], Tommy Haugen[c], Sindre M. Dyrstad[a,d]

[a]Department of Education and Sport Science, University of Stavanger, 4036 Stavanger, Norway. [b]Department of Sports Medicine, Norwegian School of Sports Sciences, 0806 Oslo, Norway. [c]Department of Sport Science and Physical Education, University of Agder, 4604 Kristiansand, Norway. [d]Department of Public Health, University of Stavanger, 4036 Stavanger, Norway.

Acknowledgements: None

Declarations of interest: The authors declare that they have no competing interests.

Corresponding author:
Andreas Åvitsland
Department of Education
University of Stavanger
4036 Stavanger, Norway
E-mail: andreas.avitsland@uis.no

Abstract

Objectives: To investigate the association between one-year changes in three components of physical fitness and mental health status in adolescents.

Design: Participants in a cluster-randomized controlled trial were treated as one cohort with a prospective study design, involving data collection at two time points separated by one year.

Methods: Norwegian 14-15-year-olds (n = 925) had valid data for both time points. Mental health was assessed via self-report questionnaire and cardiorespiratory fitness, muscular strength and body composition were objectively measured. Change scores were calculated from the physical fitness variables and serve as independent variables in linear mixed effects models.

Results: There was no association between change in body composition and psychological difficulties. Sex, immigrant status and socioeconomic status moderated associations. Subgroup results were compatible with inverse associations between change in cardiorespiratory fitness and psychological difficulties among boys (b = -0.009; 95% CI = -0.015 to -0.003; p = .006); change in muscular strength and psychological difficulties among immigrants (b = -1.97; 95% CI = -4.03 to 0.09; p = .061). Results were also compatible with an association between change in cardiorespiratory fitness and psychological difficulties among girls in the highest socioeconomic group (b = 0.014; 95% CI = 0.003 to 0.025; p = .014).

Conclusions: Increased cardiorespiratory fitness was associated with a beneficial mental health status among boys; increased muscular strength was associated with a beneficial mental health status among immigrants; and increased CRF was associated with a disadvantageous mental health status among high SES girls. The associations for different fitness components were dependent on different moderators. Possibly, this indicates that associations in different subgroups are mediated by different mechanisms. Moderated associations should be addressed in future investigations.

1. Introduction

Mental health problems are the leading cause of disability among adolescents (Erskine et al., 2015), and because most adult mental disorders begin in adolescence (Jones, 2013), it is important to investigate viable methods of prevention. There is compelling evidence in favor of physical activity (PA) as a preventive measure against mental disorders in adults (Choi et al., 2019), however, evidence is more inconclusive among adolescents (Biddle, Ciaccioni, Thomas, & Vergeer, 2019).

It is unclear how PA influences mental health, however, Lubans et al. (2016) have suggested that the influence may be caused by different neurobiological-, psychosocial- or behavioral mechanisms. For instance, the neurobiological mechanism can refer to how aerobic PA can influence structures and functions in various regions of the brain (Matta Mello Portugal et al., 2013). Furthermore, the psychosocial mechanism can refer to how strength training can positively influence senses of "self" (e.g. self-esteem, self-efficacy, self-perceptions; Collins, Booth, Duncan, Fawkner, & Niven, 2019). Lastly, the behavioral mechanism can refer to how various aerobic PA may influence self-regulation skills, such as attention and concentration (Laberge, Bush, & Chagnon, 2012). In turn, the mechanisms might influence various mental health outcomes, such as depression (Nestler et al., 2002; Steiger, Allemand, Robins, & Fend, 2014) or well-being (Steinberg, 2014, p. 177). Knowledge of the mechanisms is "critical to developing effective targeted interventions to promote mental health and reduce mental disorders in youth" (Doré et al., 2020). Unfortunately, we still know little about the mechanisms and Biddle et al. (2019) have called on future research to focus on identifying them.

To improve our ability to identify the mechanisms, measurement precision and validity is important, considering that the mechanisms can be moderated by "frequency, intensity, time, type and context" of PA (Lubans et al., 2016). This level of detail in PA measurement can prove challenging, considering how complex and multidimensional PA behavior is (Warren et al., 2010). Physical fitness is strongly related to PA (Fogelholm, Stigman, Huisman, & Metsämuuronen, 2008) and can be divided into three health-related components:

cardiorespiratory fitness (CRF), muscular strength and body composition (Caspersen, Powell, & Christenson, 1985). Although these components do not necessarily provide specific information regarding frequency, intensity, time, type and context of PA, the components can at least give some indication of some of these moderators and how they occur over time (Blair, Cheng, & Holder, 2001). Thus, physical fitness may be a more suitable measure in the effort to investigate the association between habitual PA and mental health, and the potential explanatory mechanisms of the association. Moreover, "an important distinction between physical activity and fitness is the intraindividual day-to-day variability; physical activity will undoubtedly vary on a daily basis, whereas fitness will remain relatively static, taking time to change" (Warren et al., 2010). Thus, objectively measured health-related components of physical fitness might be a more stable measurement of habitual PA. This has also been advocated in previous studies, which have measured physical fitness either as a "surrogate measure of physical activity exposure" (Dishman et al., 2012) or "a physiological outcome achieved through prolonged physical activity" (Ruggero, Petrie, Sheinbein, Greenleaf, & Martin, 2015). Other studies that argue for using physical fitness over PA have suggested that "quantification of fitness is more objective than physical activity" (Schuch et al., 2016) and that "fitness is a more relevant disease or prognostic marker" (Tacchi, Heggelund, & Scott, 2019). It has also been suggested that physical fitness itself mediates the effect of PA on mental health outcomes (Eddolls et al., 2018; Ruggero et al., 2015).

There is an abundance of studies examining the cross-sectional association between physical fitness and mental health among adults (Papasavvas, Bonow, Alhashemi, & Micklewright, 2016; Pereira-Miranda, Costa, Queiroz, Pereira-Santos, & Santana, 2017; Volaklis, Mamadjanov, Meisinger, & Linseisen, 2019). Longitudinal studies that examine the association over time, however, are scarce. The existing evidence suggests that having a lower body mass index (BMI) and avoiding CRF from declining is prospectively associated with lower odds of depression (Bjerkeset, Romundstad, Evans, & Gunnell, 2008; Dishman et al., 2012; Shigdel, Stubbs, Sui, & Ernstsen, 2019). Regarding muscular strength, a recent study indicated that grip strength is associated with lower odds for developing depression, for females only (McDowell, Gordon, & Herring, 2018).

In the context of physical fitness and mental health research, adolescents are an understudied population. Cross-sectional studies have found positive associations between mental health outcomes and CRF (Greenleaf, Petrie, & Martin, 2010), muscular strength (Smith et al., 2014) and body composition (Eddolls et al., 2018). Longitudinal studies, however, are scarce, which

should be considered a major gap in our knowledge of how physical fitness is associated with mental health over time, in the period of adolescent development. The first and only review of such longitudinal studies was published recently by Tacchi et al. (2019). The results of the review associated higher levels of physical fitness with less onset of mental health problems; however, the authors listed several limitations regarding the reviewed studies. For instance, there was an overweight of male participants, CRF was the only fitness component included, and studies did not control for important covariates, such as socioeconomic status (SES). Examples of categorical variables that are associated with both mental health and physical fitness include SES (Bøe, Øverland, Lundervold, & Hysing, 2012; Cleland, Ball, Magnussen, Dwyer, & Venn, 2009), sex (Fröjd et al., 2008) and immigrant status (Abebe, Lien, & Hjelde, 2014; Lämmle, Worth, & Bös, 2012). It is therefore important to control for these variables and to investigate whether they have moderating effects on the association between physical fitness and mental health.

To learn more about the possible preventive properties of the health-related components of physical fitness, there is a need for more research on the prospective association between physical fitness and mental health outcomes in adolescent populations. The aim of the present paper was therefore to investigate whether one-year changes in physical fitness components were associated with mental health status among Norwegian adolescents. Secondary aims were to explore potential moderators of the association, and to explore whether an association could be attributed to specific aspects of the mental health instrument.

2. Methods

2.1 Design and participants

This was a prospective cohort study that used baseline (May-August 2017) and follow-up (April-June 2018) data from the School in Motion study. Briefly, School in Motion was a multicenter cluster randomized controlled trial (RCT) that took place in four geographical regions in Norway. The primary purpose of School in Motion was to investigate whether two separate school-based PA interventions had an effect on PA levels. The secondary purpose was to investigate the effect on physical fitness, mental health, academic achievement and learning environment. Participants were ninth grade students from 29 lower secondary schools (n = 2084; age 14-15) and the schools were randomized into two intervention groups and one control group. In the present paper, all participants are treated as one cohort, but we control for school

clustering and experimental group allocation in the analyses. Figure 1 shows participant flow and missing values between baseline and follow-up. School in Motion and its procedures were approved by the Norwegian Centre for Research Data (project number 49094), and the study is in line with the Declaration of Helsinki for experiments involving humans. The study is registered in ClinicalTrials.gov ID nr: NCT03817047.

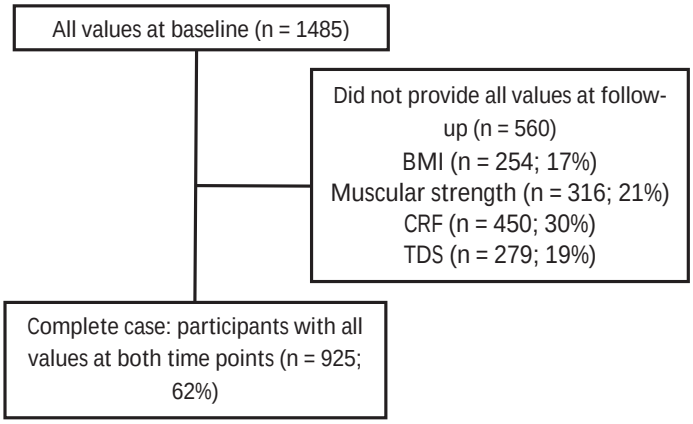


Fig. 1. Participant flow chart between baseline and follow-up. Percentages represent amount relative to 1485; BMI = body mass index; CRF = cardiorespiratory fitness; TDS = Total Difficulties Score.

2.2 Measurements

Measurement procedures for all variables have been described in detail in the paper outlining the cross-sectional associations between physical fitness and mental health at baseline (Åvitsland et al., 2020). Therefore, a brief description is provided below.

Mental health

Mental health was assessed by using the self-report version of the Strengths and difficulties questionnaire (SDQ; Goodman, 1997). The SDQ consists of 25 items, divided into five subscales, each holding five items. Four of the subscales – emotional problems, conduct problems, hyperactivity and peer problems – are summed to create the Total difficulties score (TDS). The TDS ranges from 0-40 and a higher score indicates increased psychological difficulties. The TDS functions as a dimensional measure that indicates a general mental health state in children and adolescents, as it has been found to detect changes in psychopathology on each point of the scale (Goodman & Goodman, 2009). To differentiate between groups and

identify risk of mental disorders, the TDS can be divided into three levels: Scoring 0-15 is "normal", scoring 16-19 is "borderline", and scoring 20-40 is "abnormal". The abnormal range has been shown to indicate an increased risk of developing mental disorders (Goodman, Ford, Simmons, Gatward, & Meltzer, 2000). The psychometric properties of the SDQ have been validated several times in many countries, including Norway (Van Roy, Veenstra, & Clench-Aas, 2008).

Physical fitness

We measured CRF with a modified version of the Andersen-test (Andersen, Andersen, Andersen, & Anderssen, 2008). The modification was the test field distance, which in the standardized Andersen-test is supposed to be 20 meters, but because of limited space in the gymnasiums, we used 16 meters. The outcome variable for CRF in this study is the total distance run in meters (m). To assess muscular strength, the participants performed sit-ups, standing broad jump and handgrip test, as they are described in the EUROFIT test battery (Council of Europe, 1988). As an expression of body composition, we measured participants' height and weight and calculated BMI (kg/m^2). All physical fitness tests were conducted at the participants' respective schools, by trained test-personnel.

Covariates and subgroups

We obtained categorical data for covariates and potential moderators. Sex was noted by test-personnel. Immigrant status was self-reported in the questionnaire by asking if the participant was born in Norway. Socioeconomic status (SES), is expressed as the parent with the highest education level (Erola, Jalonen, & Lehti, 2016), which we obtained from national registries. The education level was split in four subgroups: lower secondary school or less; upper secondary school; less than four years university education; more than four years university education. Based on the "normal", "borderline" and "abnormal" TDS cutoffs, we created a categorical baseline TDS variable. Lastly, we also used the categorical data describing participants' allocation to one of the three experimental groups. To ensure the credibility of our subgroup analyses, we adhere to the three most critical criteria set forth by Sun et al. (2012): subgroup variables must be assessed at baseline, subgroup hypotheses must be specified ahead of analyses and there must be an interaction effect.

2.3 Statistical analyses

All analyses were performed in IBM SPSS Statistics 25 (IBM, Armonk, New York, USA). SDQ data were managed and organized into the predetermined scales by the syntax provided by the SDQ information web page (Youthinmind, 2018). Cronbach's alpha was employed to assess the internal consistency of TDS and its subscales. The respective alphas from baseline and follow-up were as follows: emotional problems (0.67 and 0.71), conduct problems (0.51 and 0.53), hyperactivity (0.66 and 0.68), peer problems (0.61 and 0.61) and TDS (0.62 and 0.61). We report all values from both time points with descriptive statistics as means and standard deviations (SD). Paired t-tests were conducted with all outcome variables to assess the overall changes from baseline to follow-up.

We calculated individual change scores for the physical fitness variables (follow-up - baseline), which functioned as the independent variables. The change score for muscular strength is a mean composite z-score, stratified for sex. It was constructed from the change scores for handgrip test, standing broad jump and sit-ups, which were transformed into separate z-scores and averaged.

To test the prospective association between changes in health-related components of physical fitness and mental health status, we conducted linear mixed effects models, with schools as a random effect. We assessed the potential moderating effect of sex, SES, immigrant status, TDS category at baseline (normal, borderline and abnormal) and experimental group on the association between the independent and dependent variables by creating separate models for each moderator. In each model we added the moderator as an interaction term for all three independent variables, while also controlling for the independent variables. Subsequently, we performed subgroup analyses of the categorical variables that indicated an interaction effect. Follow-up TDS was the primary dependent variable and when a model indicated results that were compatible with an association, we conducted analyses to investigate if the association could be attributed to any of the TDS subscales (emotional problems, conduct problems, hyperactivity and peer problems). All models controlled for the dependent variable's baseline result, sex, SES, immigration status and experimental group allocation. We report unstandardized regression coefficients (b) with 95% confidence intervals (CI) and exact p values (b; 95% CI; p). The b represents difference on the scale of the dependent variable for each unit increase on the scale of the independent variable.

We do not dichotomously interpret the p-values to be either significant or non-significant. Instead, we adhere to the guidelines provided by the American Statistical Association's statement on statistical significance and p-values (Wasserstein & Lazar, 2016), and more recent guidelines (Wasserstein, Schirm, & Lazar, 2019). Specifically, we interpret p-values as continuous quantities that express how compatible the observed data are with the null-hypotheses: Higher p-values indicate greater compatibility with the null-hypotheses. We interpret and present our results based on a continuous interpretation of the p-values, the size of the unstandardized regression estimates (b) and the limits of the confidence intervals (Amrhein, Greenland, & McShane, 2019; Greenland et al., 2016). Our interpretations are also influenced by prior evidence, plausibility of mechanism, study design and data quality (McShane, Gal, Gelman, Robert, & Tackett, 2019).

Missing values

Missing values between enrollment and baseline have been described extensively in a previous paper (Åvitsland et al., 2020) which argued that the cross-sectional associations at baseline were likely not affected by the missing values. Of the participants with all values at baseline (n = 1485), 62% (n = 925) had all values at follow-up. A grouping variable was created, and one-way ANOVAs were used to analyze differences between the complete case group, and participants with present values at baseline but missing values at follow-up (n = 560).

To assess whether missing values were missing completely at random (MCAR), Little's MCAR test was conducted on the group with all values at baseline (n = 1485). The MCAR test of missing values for TDS, CRF, BMI and muscular strength at follow-up indicated that the missing data in this group were either missing at random (MAR) or not missing at random (MNAR; 47.401, DF = 25, p = .004). To investigate missing data further, we used descriptive and frequency statistics to determine the degree of missing values from each variable.

As a final measure of handling missing data, we conducted multiple imputation. With five imputations and ten iterations, missing data were imputed from TDS, CRF, muscular strength and body composition at both timepoints and covariates, using the automatic procedure. SPSS cannot handle multiple imputation with too many variables, thus we did not impute on the SDQ subscales (Mustillo & Kwon, 2015). The linear mixed effects models that were compatible with effects on the complete case dataset, were conducted on the imputed dataset. We report results from complete case analyses and after multiple imputation, which has been recommended by Manly and Wells (2015) and Sterne et al. (2009).

3. Results

3.1 Descriptives and group differences

The complete case group included 925 participants, of which 474 (51%) were girls and 74 (8%) were immigrants. The four SES groups contained, from lowest to highest SES, 43 (5%), 246 (27%), 374 (40%) and 262 (28%) participants. Table 1 shows the means with standard deviations from both time points, and the respective mean changes (the mean results from baseline and follow-up respective to each categorical moderator subgroup is presented in Supplementary table 1a). All variables, except conduct problems and peer problems indicated increases at follow-up.

Table 1. Mean results and difference between baseline and follow-up

	Baseline (SD)	Follow-up (SD)	Mean change	95% CI	p
Total difficulties score (0-40)	10.1 (5.0)	10.6 (5.4)	0.5	0.2 to 0.8	.001
Emotional problems (0-10)	3.0 (2.2)	3.2 (2.4)	0.2	0.1 to 0.4	<.001
Conduct problems (0-10)	1.5 (1.4)	1.6 (1.5)	0.03	0.07 to 0.1	.535
Hyperactivity (0-10)	3.9 (2.1)	4.1 (2.2)	0.2	0.1 to 0.3	.001
Peer problems (0-10)	1.7 (1.6)	1.8 (1.7)	0.1	0.1 to 0.2	.371
Body mass index	19.8 (2.9)	20.3 (3.1)	0.5	0.5 to 0.6	<.001
Standing broad jump (cm)	174 (25)	181 (26)	7	6 to 8	<.001
Sit-ups (n/30 seconds)	19 (4)	20 (4)	1	0.8 to 1.2	<.001
Handgrip strength (kg)	30.5 (7)	33.3 (9)	2.8	2.4 to 3.1	<.001
Cardiorespiratory fitness (m)	909 (91)	927 (101)	18	13 to 22	<.001

3.2 Associations between change in fitness and mental health status

The results were compatible with an inverse association between change in CRF and follow-up TDS for all participants (b = -0.004; 95% CI = -0.008 to -0.001; p = .040). The results were not compatible with associations between follow-up TDS and change in BMI or change in muscular strength. The results from moderator analyses were compatible with three separate interaction effects for associations with follow-up TDS: sex * change CRF (p = .010); immigration status * change muscular strength (p = .110); and SES * change CRF (p = .011).

Subsequent subgroup analysis of sexes was compatible with an inverse association between change in CRF and follow-up TDS among boys only (-0.009; 95% CI = -0.015 to -0.003; p =

.006). These results suggest that 100m increase in CRF was associated with 8.5% (0.9 points) lower follow-up TDS, relative to boys' mean values. (10.5 points). Subscale analyses showed that the results could mainly be attributed to lower conduct problems (-0.002; 95% CI = -0.005 to -0.001; p = .034), hyperactivity (-0.003; 95% CI = -0.006 to -0.001; p = .011) and peer problems (-0.003; 95% CI = -0.005 to -0.001; p = .005).

The subgroup analysis of immigrants/non-immigrants was compatible with an inverse association between change in strength and TDS among immigrants only (-1.96; 95% CI = -4.03 to 0.092; p = .061). These results suggest that 1 standard deviation increase in the z-score composite of handgrip strength, standing broad jump and sit-ups, was associated with 18% (1.96 points) lower follow-up TDS, relative to the immigrant population's mean values (10.7). Subscale analyses showed that the results could mainly be attributed to lower conduct problems (-0.48; 95% CI = -1.05 to 0.08; p = .095), and hyperactivity (-0.64; 95% CI = -1.45 to 0.16; p = .113).

The subgroup analysis of SES indicated inverse associations between change in CRF and follow-up TDS in SES groups 1-3 (unstandardized coefficients between -0.007 and -0.04, p-values between .039 and .063). SES group 4 (highest levels of parental education) demonstrated an opposite tendency (b = 0.004; 95% CI = -0.004 to 0.012; p = .278). This prompted further investigation by stratifying the SES groups. The analysis stratified for SES and sex was compatible with an association between change in CRF and follow-up TDS among girls in SES group 4 (n = 135; 0.014; 95% CI = 0.003 to 0.025; p = .014). These results suggest that 100m increase in CRF was associated with 13% (1.4 point) higher follow-up TDS, relative to girls in SES group 4's mean values (10.5). Subscale analyses showed that these results could mainly be attributed to emotional problems (0.005; 95% CI = -0.001 to 0.01; p = .076) and hyperactivity (0.007; 95% CI = 0.003 to 0.012; p = .002). No other sub-stratification of SES showed results compatible with associations that have not been described. Results from the full study sample and subgroups are presented in Table 2.

Table 2.
Associations between independent variables (change scores T2-T1) and dependent variables at T2.

Independent variable	Group	Dependent variable	Unstandardized b (95% CI)	p
Change CRF		**Follow-up TDS**		
	All participants		-0.004 (-0.008 to -0.001)	.040
	Boys		-0.009 (-0.015 to -0.003)	.006
	Girls		0.001 (-0.004 to 0.006)	.739
	Girls in SES group 1		-0.038 (-0.106 to 0.029)	.238
	Girls in SES group 2		-0.004 (-0.015 to 0.007)	.473
	Girls in SES group 3		-0.001 (-0.008 to 0.005)	.704
	Girls in SES group 4		0.014 (0.003 to 0.025)	.014
		Emotional problems		
	Girls in SES group 4		0.005 (-0.001 to 0.01)	.076
		Conduct problems		
	Boys		-0.002 (-0.005 to -0.001	.034
		Hyperactivity		
	Boys		-0.003 (-0.006 to -0.001)	.011
	Girls in SES group 4		0.007 (0.003 to 0.012)	.002
		Peer problems		
	Boys		-0.003 (-0.005 to -0.001)	.005
Change strength		**Follow-up TDS**		
	All participants		-0.14 (-0.59 to 0.30	.523
	Immigrants		-1.97 (-4.03 to 0.09)	.061
	Non-immigrants		-0.05 (-0.50 to 0.40)	.830
		Conduct problems		
	Immigrants		-0.48 (-1.05 to 0.086)	.095
		Hyperactivity		
	Immigrants		-0.645 (-1.45 to 0.157)	.113
Change BMI		**TDS**		
	All participants		-0.10 (-0.33 to 0.13)	.400

Note. All models included the three physical fitness variables, experimental group belonging, the dependent variable at baseline, sex, immigration and socioeconomic status, except in models where the categorical variable itself stratified the data. b = regression coefficient; CI = confidence interval; CRF = Cardiorespiratory fitness; BMI = Body mass index; SES = Socioeconomic status

3.3 Missing value analyses

Missing value analyses showed that results were compatible with some differences between the complete case group and the group with missing values (n = 560). Specifically, the complete case group displayed 6% higher SES ($p < .001$); 6% lower baseline TDS (0.6 points; $p = .017$) and 10% lower follow-up TDS (1.1 points; $p = .002$) and ran 1% (10m; $p = .064$) farther at baseline, compared to the missing values group. The results were not compatible with any other differences. Further exploration revealed that the missingness was largely due to participants not performing the CRF-test. Between test centers, not participating in the CRF-test at follow-up caused 62% to 89% of the missing values from the follow-up data collection.

After multiple imputation, linear mixed effects models showed similar associations as in the complete case results: Change in CRF was inversely associated with follow-up TDS for all participants (n = 2044; -0.005; -0.009 to 0.001; p = .062) and subgroup analyses showed the association was only present among boys (b = -0.007; -0.013 to -0.001; p = .047). However, compared to the complete case results, results after multiple imputation were much less compatible with associations between change in muscular strength and follow-up TDS among immigrants (n = 190; -0.69; -1.95 to 0.57; p = .278); and between change in CRF and follow-up TDS among girls in SES group 4 (b = 0.002; -0.006 to 0.01; p = .588). Considering that values were imputed for 55% (n = 1119) of the study sample and that an independent variable, follow-up CRF, was the main reason for missing values, the results from the multiply imputed dataset should be treated with caution (Hughes, Heron, Sterne, & Tilling, 2019). We will therefore focus our attention on the complete case results in the discussion, although we acknowledge that our results are uncertain.

4. Discussion

We aimed to investigate whether one-year changes in physical fitness components were associated with adolescent mental health status. The main findings were 1) an inverse association between change in CRF and follow-up psychological difficulties among boys; 2) an inverse association between change in muscular strength and follow-up psychological difficulties among immigrants; and 3) an association between change in CRF and psychological difficulties among girls in the highest SES group. The three findings are discussed separately.

4.1 Boys

Based on interpretation of the p-values alone, the inverse association between change in CRF and follow-up psychological difficulties among boys appear quite certain. To interpret the meaningfulness of the association, we use the paper by Goodman and Goodman (2009), which showed that a one-point reduction in TDS has been associated with a 16% to 23% lower likelihood of developing a mental disorder. Thus, the association may be meaningful in a clinical sense, considering that psychological difficulties in boys overall increased, while 100m increase in CRF suggested 14% to 3% lower levels than the boys' mean (10.5), according to the confidence intervals.

These results are in support of the baseline findings from the same population, which were incompatible with any associations between muscular strength and mental health, or BMI and mental health (Åvitsland et al., 2020). The results are also supported by previous cross-sectional studies that only found CRF to be associated with a mental health outcome, while controlling for at least one other measure of physical fitness (Andersen et al., 2017; Rieck, Jackson, Martin, Petrie, & Greenleaf, 2013; Yeatts, Martin, & Petrie, 2017). To the best of our knowledge, only Ruggero et al. (2015) have previously investigated the prospective association between physical fitness and mental health outcomes in a similar adolescent population. However, the present results stand in contrast to the results by Ruggero et al. (2015), which showed that baseline CRF was inversely associated with follow-up depression levels among girls, not boys. Similarly, in the cross-sectional study by Greenleaf et al. (2010), higher levels of CRF indicated lower levels of depression among girls only. Moreover, a recent cross-sectional study by Janssen et al. (2020), showed stronger associations between CRF and internalizing problems among girls than among boys. Rieck et al. (2013), however, had similar findings as ours and showed that boys, but not girls, with low CRF had higher odds of elevated depression than boys classified as having high CRF. Studies have also investigated whether the effect of PA interventions on mental health is moderated by sex. A meta-analysis found that the effect on mental health outcomes was larger for boys than girls in randomized studies; the effect was opposite in non-randomized studies (Ahn & Fedewa, 2011).

Considering that adult women and men show similar favorable associations between CRF and mental health (Sui et al., 2009), the different associations in the present results may be specific to the adolescent age group. The present subscale results might provide some explanation. Firstly, conduct problems, one of the three aspects of psychological difficulties that the inverse association with TDS was attributed to, is more common among boys than girls (Button et al., 2007). This was also true for the present study population (data not shown). Lower conduct problems may be caused by an increase in self-regulation, a behavioral mechanism (Lubans et al., 2016), which has been associated with PA among adolescents (Wills, Isasi, Mendoza, & Ainette, 2007) and is more common among girls than boys (Raffaelli, Crockett, & Shen, 2005). Therefore, the potential for change in conduct problems and self-regulation is larger among boys than girls. This was demonstrated in the study by Lakes and Hoyt (2004), which also found reductions in conduct problems among boys only, and that the reduction was associated with an increase in self-regulatory skills, as a result of a martial arts program.

The inverse association between change in CRF and follow-up psychological difficulties was also attributed to hyperactivity. This may have occurred through a neurobiological mechanism (Lubans et al., 2016). According to Gapin, Labban, and Etnier (2011), there are many potential neurobiological ways that PA can influence hyperactivity, for instance by increasing blood flow to the frontal region of the brain, or by increasing the availability of dopamine and norepinephrine. This was supported in the recent review by Ng, Ho, Chan, Yong, and Yeo (2017), which concluded that "…moderately-to-intense aerobic exercise, is a beneficial and well-tolerated intervention for children and adolescents with ADHD". Furthermore, that aerobic exercise may affect hyperactivity differently for boys and girls has been observed previously (Tantillo, Kesick, Hynd, & Dishman, 2002). Unfortunately, the majority of studies investigating the effect PA may have on hyperactivity has only included a male population, which makes it difficult to evaluate potentially different associations between boys and girls (Kamp, Sperlich, & Holmberg, 2014).

Lastly, the inverse association between change in CRF and follow-up psychological difficulties among boys was also attributed to peer problems. Similar findings include Lamb and Gulliford (2011), who found lower levels of peer problems in children after an aerobic exercise program; and Sagatun, Søgaard, Bjertness, Selmer, and Heyerdahl (2007), who found an inverse association between time spent in PA and peer problems among boys. A change in peer problems may occur through a psychosocial mechanism, which specifically involves social interaction and relatedness (Lubans et al., 2016). Participation in team sports provides much opportunity for social interaction and the development of social skills (Eime, Young, Harvey, Charity, & Payne, 2013) and is therefore a potential explanation for how aerobic PA affects peer problems through the psychosocial mechanism. Lower peer problems may also be directly connected to reductions in hyperactivity and conduct problems, as adolescents who have problems in peer relationships commonly also display signs of hyperactivity (Bagwell, Molina, Pelham, & Hoza, 2001) or conduct problems (Woodward & Fergusson, 1999). The difference between boys and girls may be explained by previous results showing that, compared to girls, boys report more social support, perceived benefits, self-efficacy and fun from PA (Cardon et al., 2005).

4.2 Immigrant adolescents

The inverse association between change in muscular strength and follow-up psychological difficulties among immigrants is the least certain of the present findings, based on p-value and

confidence interval interpretation. The p-value of the association is non-significant in the traditional sense, however, it still indicates that the data conforms more toward the association hypothesis, than the null hypothesis (Greenland et al., 2016). Furthermore, the confidence interval crosses the null value, but the limits indicate that the association with 1 SD strength increase spanned from 38% lower- to 0.9% higher follow-up TDS than the immigrant means (10.7; see Supplementary Table 1a). The upper limit is not very meaningful, while the lower limit represents an association that may be meaningful in a clinical sense (Goodman & Goodman, 2009).

The inverse association between change in muscular strength and psychological difficulties is contrary to previous cross-sectional studies, which ruled out muscular strength as being associated with mental health outcomes, when controlling for CRF (Andersen et al., 2017; Kelly et al., 2010; Rieck et al., 2013; Yeatts et al., 2017). These studies, however, did not examine whether being born in the country or having immigrated to the country, moderated the associations. Given the scarcity of research about the association between physical fitness and mental health among immigrant adolescents, we can only speculate why increased muscular strength was associated with follow-up psychological difficulties in the subgroup. The association was attributed to conduct problems and hyperactivity, which points to self-esteem as a possible explanation: A review by Smith et al. (2014) showed consistent cross-sectional associations between muscular strength and perceived physical appearance, perceived sports competence, self-worth and self-esteem. These findings have also been supported in more recent research (Collins et al., 2019). Lower self-esteem has been associated with both conduct problems (Ha, Petersen, & Sharp, 2008) and hyperactivity (Edbom, Lichtenstein, Granlund, & Larsson, 2006) and it is possible that increasing muscular strength leads to improved self-esteem, which in turn leads to less conduct problems and hyperactivity. The reason why this association was present among immigrants alone may be due to immigrants having lower self-esteem than their non-immigrant peers (Bankston & Zhou, 2002); however, we do not know whether this was true for our population. Although it is an interesting and important topic, to discuss reasons why immigrant adolescents have poorer self-esteem than their non-immigrant peers, is beyond the scope of this paper. However, this subgroup can experience racial discrimination, which can influence self-esteem-related outcomes, such as perceived physical appearance, feelings of belonging to a peer group and identity development (Virta, Sam, & Westin, 2004).

Importantly, these results are based on low-powered subgroup analyses and might be spurious; irrespectively, the associations should be investigated in future studies, as immigrant adolescents have been severely neglected in this research field. Accumulated knowledge is important and can be applied by persons working with integrating and acculturating immigrant adolescents.

4.3 High socioeconomic status girls

The association between change in CRF and follow-up psychological difficulties among girls in the highest SES group was surprising. The p-value suggests very low compatibility with a null-hypothesis and the upper limit of the confidence interval suggests a concerning association: 100m increase in CRF was associated with up to 24% higher follow-up psychological difficulties than the mean levels in this subgroup (10.5).

To the best of our knowledge, there is no previous research to support this association, which contradicts what we consider to be established: Aerobic PA, which is positive for CRF, is associated with improved mental health (Bailey, Hetrick, Rosenbaum, Purcell, & Parker, 2017). The association between change in CRF and psychological difficulties in the present study was attributed to hyperactivity and emotional problems. This contradicts research showing that aerobic PA has a positive effect on hyperactivity and ADHD (Ng et al., 2017), and research showing that meeting the recommended PA levels is associated with fewer emotional problems (Wiles et al., 2008). On one hand, these incongruences increase the likelihood of a spurious association. On the other hand, there may be an explanation: According to the annual report on the health and well-being of Norwegian adolescents (Bakken, 2019), stress and pressure is more often experienced by girls than boys and is also more prevalent in high SES populations. Stress has been shown to be associated with hyperactivity (Biederman et al., 1995) and emotional problems (Moksnes, Moljord, Espnes, & Byrne, 2010). Therefore, it can be hypothesized that within a female high SES group, CRF might increase as a consequence of stress and pressure from high self-expectations to exercise more often than average, and to obtain body image ideals. Furthermore, the same stress and pressure might also lead to an increase in psychological difficulties. Future studies that examine the association between change in physical fitness and mental health status among adolescents should include SES, to be able to substantiate, or contradict these findings.

4.4 Strengths and limitations

Strengths of the present study include a high number of participants, objective measurements of three components of physical fitness, adjusting for covariates, and the option to conduct interaction analyses and subsequent subgroup analyses. Importantly, subgroup analyses are debated as they are susceptible to arbitrary findings due to chance. However, the present subgroup findings can be considered credible, as we have fulfilled most of the credibility criteria for subgroup analyses, set by Sun et al. (2012). A strength of the SDQ is that it allows investigation into subscale associations, which can help us understand the underlying mechanisms that explain the relationship between PA or physical fitness and mental health.

The most important limitation of the present study is the large amount of missing data at both time points, which may have influenced the results. Additionally, the prospective cohort study design limits causal interpretations, and a longer study period with more tests could have provided more knowledge in that regard. Compared to the full study sample, the subgroups were relatively small, thus reducing the statistical power of the analyses. Furthermore, the present subgroups may not represent equivalent subgroups outside of the study population. Lastly, the SDQ is a self-report questionnaire, which can be considered a limitation because respondents might answer in ways they perceive to be socially desirable (Leising, Locke, Kurzius, & Zimmermann, 2016). Moreover, TDS and subscales performed poorly on the Cronbach's alpha-test, compared to the recommended cutoff point at 0.7 (Bland & Altman, 1997). This could be due to poor understanding of the questions, unwillingness to answer the questions honestly, or actual low consistency. This is another factor that may contribute uncertainty to the results, however, the relevance of Cronbach's alpha has been debated (Sijtsma, 2009). The present study could also have benefitted from using more than one measure of mental health.

5. Conclusions

This one-year prospective investigation has shown three separate associations indicating that increased CRF is associated with a beneficial mental health status among boys; increased muscular strength is associated with a beneficial mental health status among immigrants; and increased CRF is associated with a disadvantageous mental health status among high SES girls. The hypothesized mechanisms of the associations are plausible, and the results suggest that changes in physical fitness components influence mental health through different mechanisms,

depending on sex, immigrant status and SES. However, more research is needed on this subject, due to a scarcity of comparable studies and lack of knowledge, and the high amount of missing data, which makes the present results uncertain. Further exploration of the potentially moderated associations is necessary to improve our understanding of how we can use PA to prevent mental health problems in different subgroups.

Literature

Abebe, D. S., Lien, L., & Hjelde, K. H. (2014). What We Know and Don't Know About Mental Health Problems Among Immigrants in Norway. *Journal of Immigrant Minority Health, 16*(1), 60-67. doi:10.1007/s10903-012-9745-9

Ahn, S., & Fedewa, A. L. (2011). A Meta-analysis of the Relationship Between Children's Physical Activity and Mental Health. *Journal of Pediatric Psychology, 36*, 385-397. doi:10.1093/jpepsy/jsq107

Amrhein, V., Greenland, S., & McShane, B. (2019). Scientists rise up against statistical significance. *Nature, 567*, 305-307. doi:10.1038/d41586-019-00857-9

Andersen, J. R., Natvig, G. K., Aadland, E., Moe, V. F., Kolotkin, R. L., Anderssen, S. A., & Resaland, G. K. (2017). Associations between health-related quality of life, cardiorespiratory fitness, muscle strength, physical activity and waist circumference in 10-year-old children: the ASK study. *Quality of Life Research, 26*(12), 3421-3428. doi:10.1007/s11136-017-1634-1

Andersen, L., Andersen, T. E., Andersen, E., & Anderssen, S. (2008). An intermittent running test to estimate maximal oxygen uptake: the Andersen test. *Journal of Sports Medicine and Physical Fitness, 48*(4), 434-437. Retrieved from https://search.proquest.com/docview/202677482?accountid=136945

Bagwell, C. L., Molina, B. S. G., Pelham, W. E., & Hoza, B. (2001). Attention-Deficit Hyperactivity Disorder and Problems in Peer Relations: Predictions From Childhood to Adolescence. *Journal of the American Academy of Child & Adolescent Psychiatry, 40*(11), 1285-1292. doi:https://doi.org/10.1097/00004583-200111000-00008

Bailey, A. P., Hetrick, S. E., Rosenbaum, S., Purcell, R., & Parker, A. G. (2017). Treating depression with physical activity in adolescents and young adults: a systematic review and meta-analysis of randomised controlled trials. *Psychol Med, 48*(7), 1-20. doi:10.1017/S0033291717002653

Bakken, A. (2019). *Ungdata. Nasjonale resultater 2019 [Youth data. National results]*. Retrieved from Oslo: https://fagarkivet.oslomet.no/en/item/ungdata-2019-nasjonale-resultater [Norwegian]

Bankston, C. L., III, & Zhou, M. (2002). Being well vs. doing well: Self-esteem and school performance among immigrant and nonimmigrant racial and ethnic groups. *The International Migration Review, 36*(2), 389-415. doi:http://dx.doi.org/10.1111/j.1747-7379.2002.tb00086.x

Biddle, S. J. H., Ciaccioni, S., Thomas, G., & Vergeer, I. (2019). Physical activity and mental health in children and adolescents: An updated review of reviews and an analysis of causality. *Psychology of Sport and Exercise, 42*, 146-155. doi:https://doi.org/10.1016/j.psychsport.2018.08.011

Biederman, J., Milberger, S., Faraone, S. V., Kiely, K., Guite, J., Mick, E., . . . Reed, E. (1995). Family-Environment Risk Factors for Attention-Deficit Hyperactivity Disorder: A Test of Rutter's Indicators of Adversity. *Archives of General Psychiatry, 52*(6), 464-470. doi:10.1001/archpsyc.1995.03950180050007

Bjerkeset, O., Romundstad, P., Evans, J., & Gunnell, D. (2008). Association of Adult Body Mass Index and Height with Anxiety, Depression, and Suicide in the General PopulationThe HUNT Study. *American Journal of Epidemiology, 167*(2), 193-202. doi:10.1093/aje/kwm280

Blair, S. N., Cheng, Y., & Holder, J. S. (2001). Is physical activity or physical fitness more important in defining health benefits? *Medicine & Science in Sports & Exercise, 33*(6, Suppl.), 379-399. doi:10.1097/00005768-200106001-00007.

Bland, J. M., & Altman, D. G. (1997). Statistics notes: Cronbach's alpha. *BMJ, 314*(7080), 572. doi:10.1136/bmj.314.7080.572

Button, T. M. M., Corley, R. P., Rhee, S. H., Hewitt, J. K., Young, S. E., & Stallings, M. C. (2007). Delinquent peer affiliation and conduct problems: A twin study. *Journal of Abnormal Psychology, 116*(3), 554-564. doi:10.1037/0021-843X.116.3.554

Bøe, T., Øverland, S., Lundervold, A. J., & Hysing, M. (2012). Socioeconomic status and children's mental health: results from the Bergen Child Study. *Social Psychiatry and Psychiatric Epidemiology, 47*(10), 1557-1566. doi:10.1007/s00127-011-0462-9

Cardon, G., Philippaerts, R., Lefevre, J., Matton, L., Wijndaele, K., Balduck, A.-L., & Bourdeaudhuij, I. (2005). Physical Activity Levels In 10 To 11 Year-olds: Clustering Of Psycho-social Correlates. *Public Health Nutrition, 8*, 896-903. doi:10.1079/PHN2005738

Caspersen, C. J., Powell, K. E., & Christenson, G. M. (1985). Physical activity, exercise and physical fitness: Definitions and distinctions for health-related research. *Public Health Reports, 100*(2), 126-130. Retrieved from http://europepmc.org/backend/ptpmcrender.fcgi?accid=PMC1424733&blobtype=pdf

Choi, K. W., Chen, C.-Y., Stein, M. B., Klimentidis, Y. C., Wang, M.-J., Koenen, K. C., & Smoller, J. W. (2019). Assessment of bidirectional relationships between physical activity and depression among adults: A 2-sample mendelian randomization study. *JAMA Psychiatry, 76*(4), 399-408. doi:10.1001/jamapsychiatry.2018.4175

Cleland, V. J., Ball, K., Magnussen, C., Dwyer, T., & Venn, A. (2009). Socioeconomic Position and the Tracking of Physical Activity and Cardiorespiratory Fitness From Childhood to Adulthood. *American Journal of Epidemiology, 170*(9), 1069-1077. doi:10.1093/aje/kwp271

Collins, H., Booth, J., Duncan, A., Fawkner, S., & Niven, A. (2019). The Effect of Resistance Training Interventions on 'The Self' in Youth: a Systematic Review and Meta-analysis. *Sports Medicine - Open, 5*. doi:10.1186/s40798-019-0205-0

Dishman, R. K., Sui, X., Church, T. S., Hand, G. A., Trivedi, M. H., & Blair, S. N. (2012). Decline in Cardiorespiratory Fitness and Odds of Incident Depression. *American Journal of Preventive Medicine, 43*(4), 361-368. doi:10.1016/j.amepre.2012.06.011

Doré, I., Sylvester, B., Sabiston, C., Sylvestre, M.-P., O'Loughlin, J., Brunet, J., & Bélanger, M. (2020). Mechanisms underpinning the association between physical activity and mental health in adolescence: a 6-year study. *International Journal of Behavioral Nutrition and Physical Activity, 17*(1), 9. doi:10.1186/s12966-020-0911-5

Edbom, T., Lichtenstein, P., Granlund, M., & Larsson, J.-O. (2006). Long-term relationships between symptoms of Attention Deficit Hyperactivity Disorder and self-esteem in a prospective longitudinal study of twins. *Acta Paediatrica, 95*(6), 650-657. doi:10.1111/j.1651-2227.2006.tb02311.x

Eddolls, W. T. B., McNarry, M. A., Lester, L., Winn, C. O. N., Stratton, G., & Mackintosh, K. A. (2018). The association between physical activity, fitness and body mass index on mental well-being and quality of life in adolescents. *Quality of Life Research, 27*(9), 2313-2320. doi:10.1007/s11136-018-1915-3

Eime, R. M., Young, J. A., Harvey, J. T., Charity, M. J., & Payne, W. R. (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), 98. doi:10.1186/1479-5868-10-98

Erola, J., Jalonen, S., & Lehti, H. (2016). Parental education, class and income over early life course and children's achievement. *Research in Social Stratification and Mobility, 44*, 33-43. doi:https://doi.org/10.1016/j.rssm.2016.01.003

Erskine, H. E., Moffitt, T. E., Copeland, W. E., Costello, E. J., Ferrari, A. J., Patton, G., . . . Scott, J. G. (2015). A heavy burden on young minds: the global burden of mental and substance use disorders in children and youth. *Psychol Med, 45*(7), 1551-1563. doi:10.1017/s0033291714002888

Council of Europe. (1988). *Handbook of the Eurofit tests of Physical Fitness*. Rome: Italian National Olympic Committee.

Fogelholm, M., Stigman, S., Huisman, T., & Metsämuuronen, J. (2008). Physical fitness in adolescents with normal weight and overweight. *Scandinavian journal of medicine & science in sports, 18*(2), 162-170. doi:10.1111/j.1600-0838.2007.00685.x

Fröjd, S. A., Nissinen, E. S., Pelkonen, M. U. I., Marttunen, M. J., Koivisto, A.-M., & Kaltiala-Heino, R. (2008). Depression and School Performance in Middle Adolescent Boys and Girls. *Journal of Adolescence, 31*(4), 485-498. doi:10.1016/j.adolescence.2007.08.006

Gapin, J. I., Labban, J. D., & Etnier, J. L. (2011). The effects of physical activity on attention deficit hyperactivity disorder symptoms: The evidence. *Preventive Medicine, 52*, S70-S74. doi:https://doi.org/10.1016/j.ypmed.2011.01.022

Goodman, A., & Goodman, R. (2009). Strengths and Difficulties Questionnaire as a Dimensional Measure of Child Mental Health. *Journal of the American Academy of Child & Adolescent Psychiatry, 48*(4), 400-403. doi:https://doi.org/10.1097/CHI.0b013e3181985068

Goodman, R. (1997). The Strengths and Difficulties Questionnaire: A Research Note. *Journal of Child Psychology and Psychiatry, 38*(5), 581-586. doi:10.1111/j.1469-7610.1997.tb01545.x

Goodman, R., Ford, T., Simmons, H., Gatward, R., & Meltzer, H. (2000). Using the Strengths and Difficulties Questionnaire (SDQ) to screen for child psychiatric disorders in a community sample. *The British journal of psychiatry, 177*, 534. doi:https://doi.org/10.1192/bjp.177.6.534

Greenland, S., Senn, S. J., Rothman, K. J., Carlin, J. B., Poole, C., Goodman, S. N., & Altman, D. G. (2016). Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations. *European Journal of Epidemiology, 31*(4), 337-350. doi:10.1007/s10654-016-0149-3

Greenleaf, C., Petrie, T., & Martin, S. B. (2010). Psychosocial Variables Associated With Body Composition and Cardiorespiratory Fitness in Middle School Students. *Research Quarterly for Exercise and Sport, 81*, 65-74. doi:10.1080/02701367.2010.10599695

Ha, C., Petersen, N., & Sharp, C. (2008). Narcissism, self-esteem, and conduct problems. *European child & adolescent psychiatry, 17*(7), 406-413. doi:10.1007/s00787-008-0682-z

Hughes, R. A., Heron, J., Sterne, J. A. C., & Tilling, K. (2019). Accounting for missing data in statistical analyses: multiple imputation is not always the answer. *International Journal of Epidemiology, 48*(4), 1294-1304. doi:10.1093/ije/dyz032

Jones, P. B. (2013). Adult mental health disorders and their age at onset. *British Journal of Psychiatry, 202*(s54), s5-s10. doi:10.1192/bjp.bp.112.119164

Kamp, C. F., Sperlich, B., & Holmberg, H.-C. (2014). Exercise reduces the symptoms of attention-deficit/hyperactivity disorder and improves social behaviour, motor skills, strength and neuropsychological parameters. *Acta Paediatrica, 103*(7), 709-714. doi:10.1111/apa.12628

Kelly, N. R., Mazzeo, S. E., Evans, R. K., Stern, M., Thacker, L. F., Thornton, L. M., & Laver, J. H. (2010). Physical activity, fitness and psychosocial functioning of obese

adolescents. *Mental Health and Physical Activity, 4*(1), 31-37. doi:https://doi.org/10.1016/j.mhpa.2010.11.001

Laberge, S., Bush, P. L., & Chagnon, M. (2012). Effects of a Culturally Tailored Physical Activity Promotion Program on Selected Self-Regulation Skills and Attitudes in Adolescents of an Underserved, Multiethnic Milieu. *American Journal of Health Promotion, 26*(4), e105-e115. doi:10.4278/ajhp.090625-QUAN-202

Lakes, K. D., & Hoyt, W. T. (2004). Promoting self-regulation through school-based martial arts training. *Journal of Applied Developmental Psychology, 25*(3), 283-302. doi:https://doi.org/10.1016/j.appdev.2004.04.002

Lamb, D., & Gulliford, A. (2011). Physical exercise and children's self-concept of emotional and behavioural well-being: A randomised controlled trial. *Educational & Child Psychology, 28*(4), 66-74. Retrieved from https://www.researchgate.net/profile/Anthea_Gulliford/publication/265931587_Physical_exercise_and_children's_self-_concept_of_emotional_and_behavioural_well-being_A_randomised_controlled_trial_Physical_exercise_and_well-being/links/542174090cf203f155c6c02f.pdf

Leising, D., Locke, K. D., Kurzius, E., & Zimmermann, J. (2016). Quantifying the Association of Self-Enhancement Bias With Self-Ratings of Personality and Life Satisfaction. *Assessment, 23*(5), 588-602. doi:10.1177/1073191115590852

Lubans, D. R., Richards, J., Hillman, C., Faulkner, G., Beauchamp, M., Nilsson, M., . . . Biddle, S. (2016). Physical Activity for Cognitive and Mental Health in Youth: A Systematic Review of Mechanisms. *Pediatrics, 138*(3), e20161642. doi:10.1542/peds.2016-1642

Lämmle, L., Worth, A., & Bös, K. (2012). Socio-demographic correlates of physical activity and physical fitness in German children and adolescents. *European Journal of Public Health, 22*(6), 880-884. doi:10.1093/eurpub/ckr191

Manly, C. A., & Wells, R. S. (2015). Reporting the Use of Multiple Imputation for Missing Data in Higher Education Research. *Research in Higher Education, 56*(4), 397-409. doi:10.1007/s11162-014-9344-9

Matta Mello Portugal, E., Cevada, T., Sobral Monteiro-Junior, R., Teixeira Guimarães, T., da Cruz Rubini, E., Lattari, E., . . . Camaz Deslandes, A. (2013). Neuroscience of Exercise: From Neurobiology Mechanisms to Mental Health. *Neuropsychobiology, 68*(1), 1-14. doi:10.1159/000350946

McDowell, C. P., Gordon, B. R., & Herring, M. P. (2018). Sex-related differences in the association between grip strength and depression: Results from the Irish Longitudinal Study on Ageing. *Experimental Gerontology, 104*, 147-152. doi:https://doi.org/10.1016/j.exger.2018.02.010

McShane, B. B., Gal, D., Gelman, A., Robert, C., & Tackett, J. L. (2019). Abandon Statistical Significance. *The American Statistician, 73*(sup1), 235-245. doi:10.1080/00031305.2018.1527253

Moksnes, U. K., Moljord, I. E. O., Espnes, G. A., & Byrne, D. G. (2010). The association between stress and emotional states in adolescents: The role of gender and self-esteem. *Personality and Individual Differences, 49*(5), 430-435. doi:https://doi.org/10.1016/j.paid.2010.04.012

Mustillo, S., & Kwon, S. (2015). Auxiliary Variables in Multiple Imputation When Data Are Missing Not at Random. *The Journal of Mathematical Sociology, 39*(2), 73-91. doi:10.1080/0022250X.2013.877898

Nestler, E. J., Barrot, M., DiLeone, R. J., Eisch, A. J., Gold, S. J., & Monteggia, L. M. (2002). Neurobiology of Depression. *Neuron, 34*(1), 13-25. doi:https://doi.org/10.1016/S0896-6273(02)00653-0

Ng, Q. X., Ho, C. Y. X., Chan, H. W., Yong, B. Z. J., & Yeo, W.-S. (2017). Managing childhood and adolescent attention-deficit/hyperactivity disorder (ADHD) with exercise: A systematic review. *Complementary Therapies in Medicine, 34*, 123-128. doi:https://doi.org/10.1016/j.ctim.2017.08.018

Papasavvas, T., Bonow, R. O., Alhashemi, M., & Micklewright, D. J. S. M. (2016). Depression Symptom Severity and Cardiorespiratory Fitness in Healthy and Depressed Adults: A Systematic Review and Meta-Analysis. *Sports Medicine, 46*(2), 219-230. doi:10.1007/s40279-015-0409-5

Pereira-Miranda, E., Costa, P. R., Queiroz, V. A., Pereira-Santos, M., & Santana, M. L. (2017). Overweight and obesity associated with higher depression prevalence in adults: a systematic review and meta-analysis. *Journal of the American College of Nutrition, 36*(3), 223-233. doi:10.1080/07315724.2016.1261053

Raffaelli, M., Crockett, L. J., & Shen, Y.-L. (2005). Developmental Stability and Change in Self-Regulation From Childhood to Adolescence. *The Journal of Genetic Psychology, 166*(1), 54-76. doi:10.3200/GNTP.166.1.54-76

Rieck, T., Jackson, A., Martin, S. B., Petrie, T., & Greenleaf, C. (2013). Health-related Fitness, Body Mass Index, and Risk of Depression among Adolescents. *Medicine & Science in Sports & Exercise, 45*(6), 1083-1088. doi:10.1249/MSS.0b013e3182831db1

Ruggero, C. J., Petrie, T., Sheinbein, S., Greenleaf, C., & Martin, S. (2015). Cardiorespiratory Fitness May Help in Protecting Against Depression Among Middle School Adolescents. *Journal of Adolescent Health, 57*(1), 60-65. doi:https://doi.org/10.1016/j.jadohealth.2015.03.016

Sagatun, A., Søgaard, A. J., Bjertness, E., Selmer, R., & Heyerdahl, S. (2007). The association between weekly hours of physical activity and mental health: A three-year follow-up study of 15–16-year-old students in the city of Oslo, Norway. *BMC Public Health, 7*(1). doi:10.1186/1471-2458-7-155

Schuch, F. B., Vancampfort, D., Sui, X., Rosenbaum, S., Firth, J., Richards, J., . . . Stubbs, B. (2016). Are lower levels of cardiorespiratory fitness associated with incident depression? A systematic review of prospective cohort studies. *Preventive Medicine, 93*, 159-165. doi:10.1016/j.ypmed.2016.10.011

Shigdel, R., Stubbs, B., Sui, X., & Ernstsen, L. (2019). Cross-Sectional and Longitudinal Association of Non-Exercise Estimated Cardiorespiratory Fitness with Depression and Anxiety in the General Population: the HUNT Study. *Journal of Affective Disorders*. doi:https://doi.org/10.1016/j.jad.2019.04.016

Sijtsma, K. (2009). On the Use, the Misuse, and the Very Limited Usefulness of Cronbach's Alpha. *Psychometrika, 74*(1), 107-120. doi:10.1007/s11336-008-9101-0

Smith, J. J., Eather, N., Morgan, P. J., Plotnikoff, R. C., Faigenbaum, A. D., & Lubans, D. R. (2014). The Health Benefits of Muscular Fitness for Children and Adolescents: A Systematic Review and Meta-Analysis. *Sports Medicine, 44*(9), 1209-1223. doi:10.1007/s40279-014-0196-4

Steiger, A. E., Allemand, M., Robins, R. W., & Fend, H. A. (2014). Low and Decreasing Self-Esteem During Adolescence Predict Adult Depression Two Decades Later. *Journal of Personality and Social Psychology, 106*(2), 325–338. doi:10.1037/a0035133

Steinberg, L. (2014). *Age of opportunity: Lessons from the new science of adolescence*. Boston, MA: Houghton Mifflin Harcourt.

Sterne, J. A. C., White, I. R., Carlin, J. B., Spratt, M., Royston, P., Kenward, M. G., . . . Carpenter, J. R. (2009). Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *BMJ, 338*, b2393. doi:10.1136/bmj.b2393

Sui, X., Laditka, J. N., Church, T. S., Hardin, J. W., Chase, N., Davis, K., & Blair, S. N. (2009). Prospective study of cardiorespiratory fitness and depressive symptoms in women and

men. *Journal of Psychiatric Research, 43*(5), 546-552. doi:https://doi.org/10.1016/j.jpsychires.2008.08.002

Sun, X., Briel, M., Busse, J. W., You, J. J., Akl, E. A., Mejza, F., . . . Guyatt, G. H. (2012). Credibility of claims of subgroup effects in randomised controlled trials: systematic review. *BMJ, 344*, e1553. doi:10.1136/bmj.e1553

Tacchi, M. J., Heggelund, J., & Scott, J. (2019). Predictive validity of objective measures of physical fitness for the new onset of mental disorders in adolescents and young adults. *Early Intervention in Psychiatry, 13*(6), 1310-1318. doi:10.1111/eip.12783

Tantillo, M., Kesick, C. M., Hynd, G. W., & Dishman, R. K. (2002). The effects of exercise on children with attention-deficit hyperactivity disorder. *Medicine & Science in Sports & Exercise, 34*(2), 203-212. doi:https://doi.org/10.1097/00005768-200202000-00004

Van Roy, B., Veenstra, M., & Clench-Aas, J. (2008). Construct validity of the five-factor Strengths and Difficulties Questionnaire (SDQ) in pre-, early, and late adolescence. *J Child Psychol Psychiatry, 49*(12), 1304-1312. doi:10.1111/j.1469-7610.2008.01942.x

Virta, E., Sam, D. L., & Westin, C. (2004). Adolescents with Turkish background in Norway and Sweden: A comparative study of their psychological adaptation. *Scandinavian Journal of Psychology, 45*(1), 15-25. doi:10.1111/j.1467-9450.2004.00374.x

Volaklis, K., Mamadjanov, T., Meisinger, C., & Linseisen, J. (2019). Association between muscular strength and depressive symptoms. *Wiener klinische Wochenschrift, 131*(11), 255-264. doi:10.1007/s00508-019-1491-8

Warren, J. M., Ekelund, U., Besson, H., Mezzani, A., Geladas, N., & Vanhees, L. (2010). Assessment of physical activity–a review of methodologies with reference to epidemiological research: a report of the exercise physiology section of the European Association of Cardiovascular Prevention and Rehabilitation. *European Journal of Cardiovascular Prevention & Rehabilitation, 17*(2), 127-139. doi:10.1097/HJR.0b013e32832ed875.

Wasserstein, R. L., & Lazar, N. A. (2016). The ASA Statement on p-Values: Context, Process, and Purpose. *The American Statistician, 70*(2), 129-133. doi:10.1080/00031305.2016.1154108

Wasserstein, R. L., Schirm, A. L., & Lazar, N. A. (2019). Moving to a World Beyond "p < 0.05". *The American Statistician, 73*(sup1), 1-19. doi:10.1080/00031305.2019.1583913

Wiles, N. J., Jones, G. T., Haase, A. M., Lawlor, D. A., Macfarlane, G. J., & Lewis, G. (2008). Physical activity and emotional problems amongst adolescents. *Social Psychiatry and Psychiatric Epidemiology, 43*(10), 765. doi:10.1007/s00127-008-0362-9

Wills, T. A., Isasi, C. R., Mendoza, D., & Ainette, M. G. (2007). Self-Control Constructs Related to Measures of Dietary Intake and Physical Activity in Adolescents. *Journal of Adolescent Health, 41*(6), 551-558. doi:https://doi.org/10.1016/j.jadohealth.2007.06.013

Woodward, L. J., & Fergusson, D. M. (1999). Childhood Peer Relationship Problems and Psychosocial Adjustment in Late Adolescence. *Journal of Abnormal Child Psychology, 27*(1), 87-104. doi:10.1023/a:1022618608802

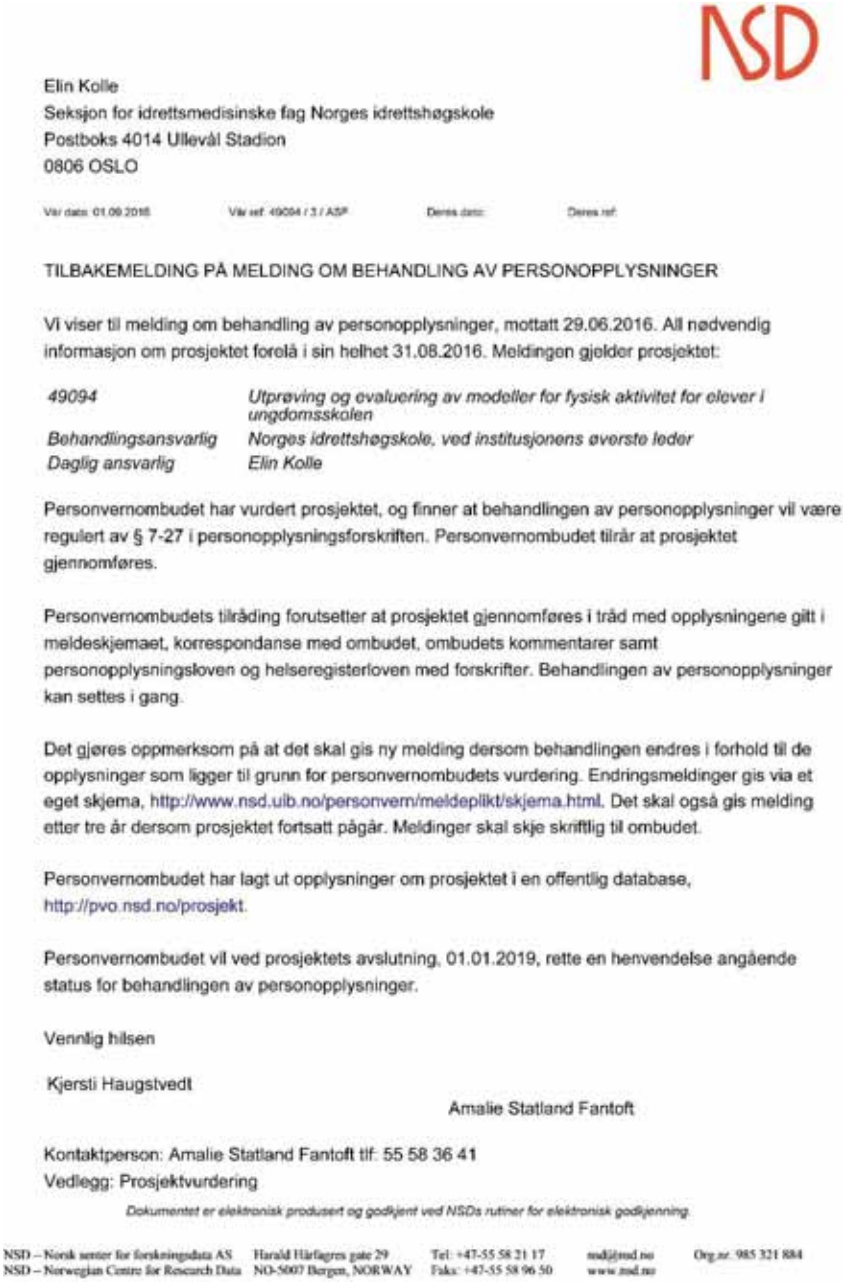
Yeatts, P. E., Martin, S. B., & Petrie, T. A. (2017). Physical fitness as a moderator of neuroticism and depression in adolescent boys and girls. *Personality and Individual Differences, 114*, 30-35. doi:10.1016/j.paid.2017.03.040

Youthinmind. (2018). Information for researchers and professionals about the Strengths & Difficulties Questionnaires. Retrieved from http://www.sdqinfo.com/

Åvitsland, A., Leibinger, E., Solberg, R., Lerum, Ø., Haugen, T., Kolle, E., & Dyrstad, S. M. (2020). The Association Between Physical Fitness and Mental Health in Norwegian Adolescents. *BMC Public Health, 20*. doi:https://doi.org/10.1186/s12889-020-08936-7

Appendices

Apppendix 1



Elin Kolle
Seksjon for idrettsmedisinske fag Norges idrettshøgskole
Postboks 4014 Ullevål Stadion
0806 OSLO

Vår dato: 01.09.2016 Vår ref: 49094 / 3 / ASF Deres dato: Deres ref:

TILBAKEMELDING PÅ MELDING OM BEHANDLING AV PERSONOPPLYSNINGER

Vi viser til melding om behandling av personopplysninger, mottatt 29.06.2016. All nødvendig informasjon om prosjektet forelå i sin helhet 31.08.2016. Meldingen gjelder prosjektet:

49094	*Utprøving og evaluering av modeller for fysisk aktivitet for elever i ungdomsskolen*
Behandlingsansvarlig	*Norges idrettshøgskole, ved institusjonens øverste leder*
Daglig ansvarlig	*Elin Kolle*

Personvernombudet har vurdert prosjektet, og finner at behandlingen av personopplysninger vil være regulert av § 7-27 i personopplysningsforskriften. Personvernombudet tilrår at prosjektet gjennomføres.

Personvernombudets tilråding forutsetter at prosjektet gjennomføres i tråd med opplysningene gitt i meldeskjemaet, korrespondanse med ombudet, ombudets kommentarer samt personopplysningsloven og helseregisterloven med forskrifter. Behandlingen av personopplysninger kan settes i gang.

Det gjøres oppmerksom på at det skal gis ny melding dersom behandlingen endres i forhold til de opplysninger som ligger til grunn for personvernombudets vurdering. Endringsmeldinger gis via et eget skjema, http://www.nsd.uib.no/personvern/meldeplikt/skjema.html. Det skal også gis melding etter tre år dersom prosjektet fortsatt pågår. Meldinger skal skje skriftlig til ombudet.

Personvernombudet har lagt ut opplysninger om prosjektet i en offentlig database, http://pvo.nsd.no/prosjekt.

Personvernombudet vil ved prosjektets avslutning, 01.01.2019, rette en henvendelse angående status for behandlingen av personopplysninger.

Vennlig hilsen

Kjersti Haugstvedt

Amalie Statland Fantoft

Kontaktperson: Amalie Statland Fantoft tlf: 55 58 36 41
Vedlegg: Prosjektvurdering

Dokumentet er elektronisk produsert og godkjent ved NSDs rutiner for elektronisk godkjenning

NSD – Norsk senter for forskningsdata AS Harald Hårfagres gate 29 Tel: +47-55 58 21 17 nsd@nsd.no Org.nr. 985 321 884
NSD – Norwegian Centre for Research Data NO-5007 Bergen, NORWAY Faks: +47-55 58 96 50 www.nsd.no

Appendix 2

Strengths and Difficulties Questionnaire

S 11-17

For each item, please mark the box for Not True, Somewhat True or Certainly True. It would help us if you answered all items as best you can even if you are not absolutely certain or the item seems daft! Please give your answers on the basis of how things have been for you over the last six months.

Your Name ..... Male/Female

Date of Birth.....

	Not True	Somewhat True	Certainly True
I try to be nice to other people. I care about their feelings	☐	☐	☐
I am restless, I cannot stay still for long	☐	☐	☐
I get a lot of headaches, stomach-aches or sickness	☐	☐	☐
I usually share with others (food, games, pens etc.)	☐	☐	☐
I get very angry and often lose my temper	☐	☐	☐
I am usually on my own. I generally play alone or keep to myself	☐	☐	☐
I usually do as I am told	☐	☐	☐
I worry a lot	☐	☐	☐
I am helpful if someone is hurt, upset or feeling ill	☐	☐	☐
I am constantly fidgeting or squirming	☐	☐	☐
I have one good friend or more	☐	☐	☐
I fight a lot. I can make other people do what I want	☐	☐	☐
I am often unhappy, down-hearted or tearful	☐	☐	☐
Other people my age generally like me	☐	☐	☐
I am easily distracted, I find it difficult to concentrate	☐	☐	☐
I am nervous in new situations. I easily lose confidence	☐	☐	☐
I am kind to younger children	☐	☐	☐
I am often accused of lying or cheating	☐	☐	☐
Other children or young people pick on me or bully me	☐	☐	☐
I often volunteer to help others (parents, teachers, children)	☐	☐	☐
I think before I do things	☐	☐	☐
I take things that are not mine from home, school or elsewhere	☐	☐	☐
I get on better with adults than with people my own age	☐	☐	☐
I have many fears, I am easily scared	☐	☐	☐
I finish the work I'm doing. My attention is good	☐	☐	☐

Do you have any other comments or concerns?

Please turn over - there are a few more questions on the other side

Overall, do you think that you have difficulties in one or more of the following areas: emotions, concentration, behaviour or being able to get on with other people?

No	Yes-minor difficulties	Yes-definite difficulties	Yes-severe difficulties
☐	☐	☐	☐

If you have answered "Yes", please answer the following questions about these difficulties:

• How long have these difficulties been present?

Less than a month	1-5 months	6-12 months	Over a year
☐	☐	☐	☐

• Do the difficulties upset or distress you?

Not at all	Only a little	Quite a lot	A great deal
☐	☐	☐	☐

• Do the difficulties interfere with your everyday life in the following areas?

	Not at all	Only a little	Quite a lot	A great deal
HOME LIFE	☐	☐	☐	☐
FRIENDSHIPS	☐	☐	☐	☐
CLASSROOM LEARNING	☐	☐	☐	☐
LEISURE ACTIVITIES	☐	☐	☐	☐

• Do the difficulties make it harder for those around you (family, friends, teachers, etc.)?

Not at all	Only a little	Quite a lot	A great deal
☐	☐	☐	☐

Your Signature .....

Today's Date .....

Thank you very much for your help

Appendix 3

Interview guide: Students model 2

The reason we are conducting this interview is that i would like to know what you think about the project that you are part of, where you get more physical activity and physical education than usual. This is not a test, there are no wrong answers. I hope you are honest and provide your true opinions, not what you think others would have answered. The interview is recorded by this recorder so that I do not have to write down all that we talk about. No names will be used and I will not tell the school, your teachers or your parents, what you tell me. The interview will last approximately 30 minutes. If you do not want to answer questions, you are allowed to, if there is a question you do not understand, let me know and I will explain better. Do you have any questions for me? Then, I will start the recorder and we will begin.

Goals/Understanding

1. Do you know that you are part of a project? Can you talk about what the project is about?
    a. Explain what they are part of if they do not know, or are unable to articulate.

Be happy – to be social in motion

1. Fidelity – to what degree does the execution correspond with the intervention's intentions:
    a. What activity have you chosen, and why?
    b. What other activity groups are in your class?
    c. Can you tell me about how the planning process occurred?
        i. How was it to choose groups?
        ii. How was it to choose activity?
        iii. What decided your choice of group, was it friends or was it the activity itself?
        iv. Are the activity groups formed within each homeroom class or across the classes?
        v. Could you decide yourselves what you wanted to do?
        vi. How much did the teacher help?
        vii. Can you tell me about how you chose a leadership structure in the group?
        viii. Can you tell me about how you worked on your group goals and activity goals?
        ix. Do you think that you have become better acquainted with students from other homeroom classes after you started the activity groups?
        x. Has it happened that you could not have BH because of tests, days off, other projects or other reasons?
2. Quality – how well has the different components been executed?
    a. Can you describe an average BH lesson?
        i. How are activities commenced?
        ii. To what extent do you receive help/guidance from your teacher during a BH lesson?
        iii. Do you ever receive help from your teacher without asking for help or guidance?
    b. Do you have access to necessary equipment to do what you want?
        i. Do you know if anyone could not pursue their chosen activities due to limited equipment?
    c. Can you tell me about where you are during the BH lessons?
    d. Can you tell me about where the other groups are during the BH lessons?
    e. Can you provide me with one or more examples of a BH lesson where you:

i. Had fun?
ii. Were social?
iii. Accomplished something difficult?
iv. Exhausted yourselves physically?

f. Can you provide me with one or more examples of a BH lesson, which:
i. Was bad, or not so fun?
1. Who were responsible for the bad lesson?

3. Responsiveness – to what extent is the program stimulating for the participants?
a. What do you think about the lessons? Elaborate as much as you can.
b. Do you think there is a difference between girls and boys regarding how much they like BH?
c. Do you think there is a difference between students who do leisure time sports and those who do not, regarding how much they like BH?
d. Can you tell me about whether you have had any arguments or conflicts during the lessons?
i. How were they solved?
ii. Do you have examples from any other groups?
e. Have there been BH lessons where you did not want to do what you were supposed to?
i. In which case, what happened?
ii. Did you do anything else?
iii. How did you get back your motivation?
f. Have any of you switched activity since the beginning?
i. Why?
ii. Have others also switched activity?
g. Do you think you have improved your skills in your chosen activity?
i. Can you give me an example?
4. Adaptation – what changes were done to make the program fit the school?
a. Adaptations will be discovered by comparing our aims for the intervention with what the students answer.

Don't worry – extra PE/social in motion.

1. Fidelity
a. Can you describe how a DW lesson usually occur?
i. Is it different from BH? In what way?
1. If you do not have BH activities, what do you do? As many examples as you can think of.
2. Are you part of choosing the activity?
3. Yes/no? What do you think about that?
4. What does the teacher do in the lessons?
ii. If you have BH activities, is the lesson identical with the BH lesson?
1. What do you think about that?
2. What do you think the other students think about that?
2. Quality
a. Do you usually have good efforts in the DW lessons?
b. Do you have enough time for activity in a single lesson?
c. Do you get help/guidance from your teacher during the lesson?
i. If yes, is it good guidance? In what way?

ii. If no, do you miss receiving guidance and help?

d. Can you give one or more examples of a DW lesson where you:
   i. Had fun?
   ii. Were social?
   iii. Accomplished something difficult?
   iv. Exhausted yourselves physically?

e. Can you give on or more examples of a DW lesson that was bad or not so fun?

3. Responsiveness
   a. What do you think about the DW lessons? Elaborate as much as you can.
   b. Do you think there is a difference between boys and girls, regarding how much they like DW?
   c. Do you think there is a difference between students who do leisure time sports and those who do not, regarding how much they like DW?
   d. Do you think an extra PE lesson has any effect on you? Good or bad?
4. Adaptation
   a. Adaptation is discovered by comparing our goals for the intervention with what the students answer.

Overall experience of the year – effect and implications

1. Can you say anything about what you think has been positive about increased physical activity this year?
2. Can you say anything about what you think has been negative about increased physical activity this year?
3. Would you like to keep having DW and BH in tenth grade?
   a. Why/why not?
4. In the project that you are involved in, researchers are trying to figure out whether adolescents participating in more physical activity in school, will improve their wellbeing in school, their physical health, mental health and their academic achievement. What do you think about this?
   a. Wellbeing and mental health?
   b. Physical health?
   c. Learning environment and social relationships?
   d. Learning and academic achievement?
5. If the government decides that all lower secondary school students will get one hour of daily physical activity in their schedule, how do you think that it should be organized?

Interview guide teachers model 2

The reason we are conducting this interview is that I would like to know what you teachers think about the project. There are no wrong answers and I am not looking for confirmation that the project is all good. I hope you are honest and that you provide your true opinions. The interview will be recorded by this recorder so I do not have to write down what you say. No names will be used and no one has to know you have said. The interview will last approximately 30 minutes. If you do not want to answer a question, that is fine. If I ask stupid questions, you are allowed to let me know. If there are any questions you do not understand, let me know and I will explain better. Do you have any questions for me? Then, I will start the recorder and we will begin.

Goals/understanding

1. What is your opinion on having more physical activity in school?
2. How was the mood among your colleagues regarding the school's participation in the study?

Be happy – being social in motion

1. Fidelity – to what extent does the execution correspond to the intervention's intentions?
    a. What activity groups are in your school?
    b. Can you tell me about how the planning process went, regarding group choices and activity choices?
        i. How did it go when the students chose groups?
        ii. How did it go when the students chose activities?
        iii. Activity groups exclusive to each homeroom classes or across classes?
        iv. What was your role in the planning process?
        v. How much did you help?
        vi. Can you tell me about how the students chose a leadership structure?
    c. Has it ever occurred that you could not have BH because of tests, days of, projects or other reasons?
    d. How many teachers are involved in the execution of BH?
        i. How is it to work together with your colleagues in these lessons?
2. Quality – how well are the different components executed?
    a. Can you describe an average BH lesson?
        i. How is the activity commenced?
        ii. To what extent do students receive help/guidance from teachers during BH?
    b. Have you had sufficient equipment to let students do what they wanted?
        i. Have any students been unable to do what they wanted because of limited equipment?
    c. Can you tell me about where the students are during BH?
        i. Do you think your presence influences the quality of the lesson? How/how not?
    d. Can you give one of more examples of a BH lesson where you, as teachers:
        i. Felt a need to intervene in what goes on?
        ii. Have been superfluous?
        iii. Have been impressed by students' ability to work from their own initiative?
        iv. Have been disappointed by students' lacking ability to work from their own initiative?
3. Responsiveness – to what extent is the program stimulating for its participants
    a. How do you think the students respond to BH? Elaborate as much as you can.
        i. How much do the teachers matter for the response?

ii. How much does the intervention matter for the response?

b. Do you think there is a difference between boys and girls, regarding how much they like BH?

c. Do you think there is a difference between those who do leisure time sports and those who do not, regarding how much they like BH?

d. Can you tell me if you have observed any arguments or conflicts in the lessons?
   i. How were they solved?
   ii. Do you have other examples that you have heard from other teachers or students?

e. Have there been BH lessons where the students did not want to do what they were supposed to?
   i. If yes, what happened?
   ii. Did they do something else?
   iii. How were the subsequent BH lessons, how did the motivation return?

f. Have students switched activity since the beginning?
   i. Why?

g. Do you think that your relationship with the students in the BH lessons is different from your relationship with the students in an ordinary PE lesson?

h. Do you think the ninth grade students in your school can handle the freedom and responsibility that comes with the intervention?
   i. Is there a difference between students regarding this?
      1. Boys and girls?
      2. Those who excel in PE and those who do not?

4. Adaptation – what changes have been done to make the program fit the school
   a. What adaptations have you done to make BH fit your school?
      i. Other adaptations that have been done along the way?
   b. Have you experienced resistance from someone, for instance the university, school leaders, parents or students, against making adaptations that you thought were necessary?

Don't worry – extra PE/being social in motion.

Questions mainly apply to teacher led DW lessons that are organized differently from BH

1. Fidelity
   a. Can you describe an average DW lesson?
      i. Is it different from BH? In what way?
         1. If you did not have BH activities, what did you do? As many examples as you can think of.
         2. Did you allow students to participate in choosing activity? Why/why not?
      ii. If you have BH activities, is the lesson identical with the BH lesson?
         1. What do you think about that?
         2. What do you think the students think about that?
2. Quality
   a. Do the students usually exhibit good efforts in the DW lessons?
   b. Do the students get to be social during the DW lessons? How/how not?
   c. Is one single lesson enough time to be in activity?
   d. What is your role in a DW lesson?

    e. In a teacher led DW, can you give one or more examples where you, as teachers:
        i. Felt the need to intervene in the lesson?
        ii. Have been superfluous?
        iii. Thought that the students acted differently than in an ordinary PE lesson?
3. Responsiveness
    a. Questions apply to teacher led DW lessons
        i. What do you think of these lessons? Elaborate as much as you can
        ii. Do you think there is a difference between boys and girls, regarding how much they like DW?
        iii. What do you think about DW compared to ordinary PE lessons? Is there a difference?
        iv. Has DW influenced how you view PE? Do you know if this has happened to any of the other teachers?
4. Adaptations
    a. What adaptations have you done to make DW fit your school?
    b. Have you experienced resistance from, for instance the university, school leaders, parents or students, against making adaptations that you thought were necessary?
    c. If you could choose freely, how would you adapt DWBH?

Overall experience of the year – effect and implications

1. Can you say something about what has been positive about extra physical activity this year?
2. Can you say something about what has been negative about extra physical activity this year?
3. Would you like to keep having DW and BH in tenth grade?
    a. Why/why not?
4. If the government decide that all lower secondary school students shall have a daily lesson physical activity in their schedule, how do you think this should be organized?

Appendix 4

Protocol for activity groups (Be happy groups)

Group name: .....

Group members

Name	Class

Rules

- *E. g.: First and foremost, we are good friends in motion – all shall be heard, all shall be seen and all shall thrive.*
- *E. g.: First and foremost, we are skiers. We shall develop each other, exercise together and get in better shape.*
- *E. g.: We do parkour. We will learn from Youtube, help each other and challenge each other.*
- *E. g.: We are homework joggers. We will run and discuss other school subjects.*
- *E. g.: We do street dance and we shall develop each other, be friends and have fun dancing.*

We have the following rules that shall govern what we are doing:

Aims for the year

- *E. g.: We will get to know each other well by participating in activity together.*
- *E. g.: We will improve as skiers so we can become trainers for the younger skiers in our organization.*
- *E. g.: We will perform a dance show at the local mall.*
- *E. g.: We shall learn so much about yoga so that we can teach others.*
- *E. g.: We will improve our English speaking skills.*

Our aim for the year is:

Short-term period aims:

- *E. g.: In the period until November first, we shall mainly have fun doing our activity.*
- *E. g.: In the period until November first, we will learn four new tricks.*
- *E. g.: In the period until November first, we will do basic training for cross country skiing.*

In the period..... our aim is to...

Long-term plans and name of the responsible person:			
Date	What we are doing	Where	Responsible for starting the lesson

Strategy for solving conflicts

E. g.: If we have disagreements that last, we have to figure out how to solve them together. We have to dare to bring up difficult subjects. If we are unable to do this, we ask help from our teacher.

If our group experience problems, we will:

1:

2:

3:

4:

Routines for reporting attendance, deviances, etc.

- *E. g.: We keep a protocol of who attends and whether things happen. Our teacher can see this whenever he or she wants to. If we experience problems, we will ask our teacher for help.*

Wes hall report attendance, deviances and other occurrences the following way:

Signatures from all group members:

Appendix 5

Supplementary Table 1a. Mean results from baseline and follow-up, and mean change between timepoints. Respective to each categorical moderator subgroup.

Subgroup	Variable	Baseline (SD)	Follow-up (SD)	Mean change	p
Boys (n = 451)	Total difficulties score (0-40)	9.9 (5)	10.5 (6.7)	0.6	.010
	Emotional problems (0-10)	2.3 (1.8)	2.5 (2.1)	0.2	.022
	Conduct problems (0-10)	1.8 (1.6)	1.9 (1.7)	0.1	.452
	Hyperactivity (0-10)	4.0 (2.1)	4.3 (2.2)	0.3	.006
	Peer problems (0-10)	1.8 (1.7)	1.9 (1.8)	0.1	.416
	Body mass index	19.3 (2.8)	20.1 (3.1)	0.7	<.001
	Standing broad jump (cm)	181 (27)	194 (27)	13	<.001
	Sit-ups (n/30 seconds)	20 (4)	21 (4)	1	<.001
	Handgrip strength (kg)	32 (8.1)	37 (9.3)	5	<.001
	Cardiorespiratory fitness (m)	934 (98)	962 (102)	28	<.001
Girls (n = 474)	Total difficulties score (0-40)	10.2 (5.1)	10.6 (5.1)	0.4	.023
	Emotional problems (0-10)	3.6 (2.8)	3.9 (2.4)	0.2	.008
	Conduct problems (0-10)	1.2 (1.3)	1.2 (1.2)	0	>.999
	Hyperactivity (0-10)	3.7 (2.1)	3.9 (2.2)	0.2	.042
	Peer problems (0-10)	1.6 (1.5)	1.6 (1.6)	0.1	.662
	Body mass index	20.2 (3)	20.6 (3)	0.4	<.001
	Standing broad jump (cm)	167 (21)	168 (21)	1	.069
	Sit-ups (n/30 seconds)	18 (4)	19 (4)	1	<.001
	Handgrip strength (kg)	29.1 (6)	29.7 (6)	0.6	.002
	Cardiorespiratory fitness (m)	886 (78)	894 (88)	8	.007
Immigrant (n = 74)	Total difficulties score (0-40)	10.2 (5.3)	10.7 (5)	0.6	.364
	Emotional problems (0-10)	3.1 (2.1)	3.1 (2.3)	0.1	.911
	Conduct problems (0-10)	1.6 (1.5)	1.6 (1.3)	0.1	.832
	Hyperactivity (0-10)	3.7 (2)	4.1 (2.1)	0.5	.042
	Peer problems (0-10)	1.8 (1.6)	2.0 (1.6)	0.2	.457
	Body mass index	19.6 (3.1)	20.1 (3.4)	0.5	<.001
	Standing broad jump (cm)	175 (23)	181 (27)	7	<.001
	Sit-ups (n/30 seconds)	19 (4)	20 (4)	1	.002
	Handgrip strength (kg)	30.4 (7)	32 (9)	1.8	.005
	Cardiorespiratory fitness (m)	903 (82)	927 (93)	24	.006
Non-immigrant (n = 851)	Total difficulties score (0-40)	10.1 (5)	10.6 (5)	0.5	.001
	Emotional problems (0-10)	2.9 (2.2)	3.2 (2.4)	0.3	<.001
	Conduct problems (0-10)	1.5 (1.4)	1.6 (1.5)	0.1	.473
	Hyperactivity (0-10)	3.9 (2.1)	4.1 (2.2)	0.2	.004
	Peer problems (0-10)	1.7 (1.6)	1.7 (1.7)	0.1	.498
	Body mass index	19.8 (3)	20.3 (3)	0.6	<.001
	Standing broad jump (cm)	174 (25)	181 (28)	7	<.001
	Sit-ups (n/30 seconds)	19 (4)	20 (4)	1	<.001
	Handgrip strength (kg)	30.6 (7)	33.4 (9)	2.8	<.001
	Cardiorespiratory fitness (m)	910 (92)	927 (101)	17	<.001
SES group 1 (n = 43)	Total difficulties score (0-40)	11.1 (5)	11.7 (6)	0.6	.509
	Emotional problems (0-10)	3.1 (2.2)	3.2 (2.4)	0.1	.717
	Conduct problems (0-10)	1.7 (1.3)	2.0 (1.9)	0.3	.313
	Hyperactivity (0-10)	4.1 (1.8)	4.5 (2.0)	0.4	.138
	Peer problems (0-10)	2.2 (1.8)	2.0 (1.6)	-0.3	.377
	Body mass index	20 (3)	20.4 (3)	0.4	.023
	Standing broad jump (cm)	171 (24)	177 (26)	6	.014
	Sit-ups (n/30 seconds)	18 (4)	19 (3)	1	.070
	Handgrip strength (kg)	29.6 (7)	32.7 (8)	3	.001
	Cardiorespiratory fitness (m)	884 (88)	908 (105)	25	.002
SES group 2 (n = 246)	Total difficulties score (0-40)	10.8 (5)	11.7 (5)	0.9	.004
	Emotional problems (0-10)	2.9 (2.1)	3.2 (2.3)	0.3	.038

	Conduct problems (0-10)	1.7 (1.6)	1.9 (1.5)	0.2	.085
	Hyperactivity (0-10)	4.4 (2.1)	4.7 (2.1)	0.3	.008
	Peer problems (0-10)	1.8 (1.6)	1.9 (1.8)	0.1	.304
	Body mass index	20.1 (4)	20.7 (4)	0.6	<.001
	Standing broad jump (cm)	174 (25)	181 (28)	6	<.001
	Sit-ups (n/30 seconds)	19 (4)	19 (4)	1	<.001
	Handgrip strength (kg)	30.8 (8)	33.5 (9)	2.7	<.001
	Cardiorespiratory fitness (m)	904 (92)	917 (103)	13	.007
SES group 3 (n = 374)	Total difficulties score (0-40)	9.9 (5)	9.9 (5)	0	.980
	Emotional problems (0-10)	2.9 (2.2)	3.1 (2.3)	0.1	.210
	Conduct problems (0-10)	1.4 (1.4)	1.4 (1.4)	0.1	.216
	Hyperactivity (0-10)	3.8 (2)	3.9 (2.2)	0.1	.599
	Peer problems (0-10)	1.7 (1.6)	1.6 (1.7)	0.1	.406
	Body mass index	20 (3)	20 (3)	0.5	<.001
	Standing broad jump (cm)	174 (25)	181 (27)	7	<.001
	Sit-ups (n/30 seconds)	19 (4)	20 (4)	1	<.001
	Handgrip strength (kg)	30.5 (7)	33.1 (8)	2.5	<.001
	Cardiorespiratory fitness (m)	912 (94)	939 (102)	18	<.001
SES group 4 (n = 262)	Total difficulties score (0-40)	9.4 (5)	10.3 (5)	0.9	.002
	Emotional problems (0-10)	3.0 (2.2)	3.4 (2.4)	0.4	.005
	Conduct problems (0-10)	1.4 (1.4)	1.5 (1.4)	0.1	.696
	Hyperactivity (0-10)	3.4 (2.1)	3.6 (2.2)	0.3	.015
	Peer problems (0-10)	1.6 (1.6)	1.8 (1.7)	0.2	.051
	Body mass index	19 (3)	20 (3)	0.5	<.001
	Standing broad jump (cm)	174 (25)	182 (27)	8	<.001
	Sit-ups (n/30 seconds)	20 (4)	20 (4)	1	<.001
	Handgrip strength (kg)	30.5 (7)	33.6 (9)	3	<.001
	Cardiorespiratory fitness (m)	914 (87)	936 (95)	22	<.001
Intervention group 1 (n = 386)	Total difficulties score (0-40)	10.3 (5)	10.6 (5)	0.3	.150
	Emotional problems (0-10)	2.9 (2.1)	3.1 (3)	0.2	.021
	Conduct problems (0-10)	1.6 (1.4)	1.5 (1.3)	0.1	.215
	Hyperactivity (0-10)	3.9 (2.1)	4.1 (2.2)	0.2	.014
	Peer problems (0-10)	1.9 (1.7)	1.8 (1.7)	0.1	.598
	Body mass index	20 (3)	20 (3)	0.6	<.001
	Standing broad jump (cm)	170 (26)	179 (26)	8	<.001
	Sit-ups (n/30 seconds)	18 (3)	20 (4)	1	<.001
	Handgrip strength (kg)	30.4 (6)	32.9 (8)	2.5	<.001
	Cardiorespiratory fitness (m)	900 (98)	931 (113)	31	<.001
Intervention group 2 (n = 230)	Total difficulties score (0-40)	10.2 (5)	10.7 (5)	0.5	.081
	Emotional problems (0-10)	3.1 (2.2)	3.2 (2.3)	0.1	.448
	Conduct problems (0-10)	1.5 (1.5)	1.7 (1.6)	0.2	.089
	Hyperactivity (0-10)	3.9 (2.2)	4.0 (2.2)	0.1	.793
	Peer problems (0-10)	1.6 (1.5)	1.8 (1.7)	0.2	.067
	Body mass index	20 (3)	20 (3)	0.6	<.001
	Standing broad jump (cm)	176 (24)	181 (29)	5	<.001
	Sit-ups (n/30 seconds)	19 (4)	20 (4)	1	<.001
	Handgrip strength (kg)	30.9 (8)	33.1 (9)	2.2	<.001
	Cardiorespiratory fitness (m)	910 (85)	912 (99)	2	.708
Control group (n = 309)	Total difficulties score (0-40)	9.7 (5)	10.4 (6)	0.8	.005
	Emotional problems (0-10)	2.9 (2.2)	3.2 (2.5)	0.3	.007
	Conduct problems (0-10)	1.4 (1.4)	1.5 (1.6)	0.1	.375
	Hyperactivity (0-10)	3.7 (2.1)	4.0 (2.2)	0.3	.005
	Peer problems (0-10)	1.6 (1.6)	1.6 (1.7)	0.1	.531
	Body mass index	20 (3)	20 (3)	0.5	<.001
	Standing broad jump (cm)	177 (23)	184 (27)	7	<.001

	Sit-ups (n/30 seconds)	20 (4)	20 (4)	1	<.001
	Handgrip strength (kg)	30.5 (7)	33.9 (9)	3.5	<.001
	Cardiorespiratory fitness (m)	920 (86)	933 (84)	13	<.001
Normal TDS at baseline (n = 799)	Total difficulties score (0-40)	8.6 (4)	9.6 (5)	1	<.001
	Emotional problems (0-10)	2.5 (1.8)	2.9 (2.2)	0.4	<.001
	Conduct problems (0-10)	1.3 (1.2)	1.4 (1.4)	0.1	.005
	Hyperactivity (0-10)	3.5 (1.9)	3.8 (2.1)	0.3	<.001
	Peer problems (0-10)	1.4 (1.3)	1.6 (1.6)	0.2	.002
	Body mass index	20 (3)	20 (3)	0.6	<.001
	Standing broad jump (cm)	174 (25)	182 (28)	7	<.001
	Sit-ups (n/30 seconds)	19 (4)	20 (4)	1	<.001
	Handgrip strength (kg)	30.7 (7)	33.3 (9)	2.6	<.001
	Cardiorespiratory fitness (m)	911 (91)	930 (100)	19	<.001
Borderline TDS at baseline (n = 84)	Total difficulties score (0-40)	17.1 (1)	15.3 (5)	-1.8	.002
	Emotional problems (0-10)	5.2 (1.9)	4.8 (2.5)	-0.4	.135
	Conduct problems (0-10)	2.8 (1.7)	2.5 (1.7)	-0.3	.117
	Hyperactivity (0-10)	6.0 (2)	5.5 (2.1)	-0.5	.033
	Peer problems (0-10)	3.1 (1.6)	2.6 (2)	-0.5	.029
	Body mass index	20 (3)	20 (3)	0.6	<.001
	Standing broad jump (cm)	170 (22)	175 (26)	15	.001
	Sit-ups (n/30 seconds)	18 (4)	19 (4)	1	.048
	Handgrip strength (kg)	29.5 (7)	32.7 (8)	3.2	<.001
	Cardiorespiratory fitness (m)	893 (95)	908 (97)	15	.061
Abnormal TDS at baseline (n = 42)	Total difficulties score (0-40)	22.9 (3)	18.9 (6)	-4	<.001
	Emotional problems (0-10)	6.9 (2)	6.0 (2.4)	-0.9	.035
	Conduct problems (0-10)	4.2 (1.6)	2.9 (1.8)	-1.3	<.001
	Hyperactivity (0-10)	6.9 (1.7)	6.2 (2.1)	-0.6	.039
	Peer problems (0-10)	4.9 (2)	3.8 (2)	-1.1	.004
	Body mass index	20 (4)	21 (4)	0.3	.179
	Standing broad jump (cm)	172 (28)	177 (30)	5	.058
	Sit-ups (n/30 seconds)	18 (4)	19 (4)	0.4	.437
	Handgrip strength (kg)	30.8 (8)	34.6 (10)	3.8	<.001
	Cardiorespiratory fitness (m)	904 (86)	912 (111)	8	.439
Girls in SES group 4 (n = 135)	Total difficulties score (0-40)	9.7 (5)	10.4 (5)	0.7	.071
	Emotional problems (0-10)	3.7 (2.3)	4.2 (2.3)	0.4	.026
	Conduct problems (0-10)	1.1 (1.2)	1.2 (1.2)	0.1	.773
	Hyperactivity (0-10)	3.3 (2.1)	3.4 (2.1)	0.1	.433
	Peer problems (0-10)	1.5 (1.4)	1.6 (1.6)	0.1	.387
	Body mass index	20 (3)	20 (3)	0.3	.001
	Standing broad jump (cm)	169 (20)	169 (21)	0.2	.884
	Sit-ups (n/30 seconds)	18 (4)	19 (4)	1	<.001
	Handgrip strength (kg)	29.5 (5)	30.3 (6)	0.8	.038
	Cardiorespiratory fitness (m)	889 (73)	906 (81)	17	.001

Note. SES = Socioeconomic status expressed as the parent with highest education. SES-group 1 = lower secondary school; SES-group 2 = upper secondary school; SES-group 3 = <4 years university education; SES-group 4 = >4years university education. TDS = Total Difficulties Score; Normal, Borderline, Abnormal TDS = 0-15, 16-19, 20-40