

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

Andreas Åvitsland^{a,*}, Eva Leibinger^a, Geir Kåre Resaland^b, Runar Barstad Solberg^c, Elin Kolle^c, Sindre M. Dyrstad^{a,d}

^a Department of Teacher Education and Sports Sciences, University of Stavanger, Norway

^b Center for Physically Active Learning, Western Norway University of Applied Sciences, Norway

^c Department of Sports Medicine, Norwegian School of Sports Sciences, Norway

^d Department of Public Health, University of Stavanger, Norway

ARTICLE INFO

Keywords:

Physical activity
Mental health
School-based intervention
Cluster randomized controlled trial

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 min per week, throughout a 29-week intervention period. M1 consisted of 30 min physically active learning, 30 min physical activity and one 60 min 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.

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%–27% (Bakken, 2015, 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

* Corresponding author.

E-mail address: Andreas.avitsland@uis.no (A. Åvitsland).

<https://doi.org/10.1016/j.mhpa.2020.100348>

Received 12 May 2020; Received in revised form 15 July 2020; Accepted 21 July 2020

Available online 1 August 2020

1755-2966/© 2020 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

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 min 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, Smith, 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, & Sogaard, 2008; G. K.; 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, Smith, 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](https://clinicaltrials.gov) ID nr: NCT03817047. Fig. 1 shows the participant flow from enrollment to post-testing.

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

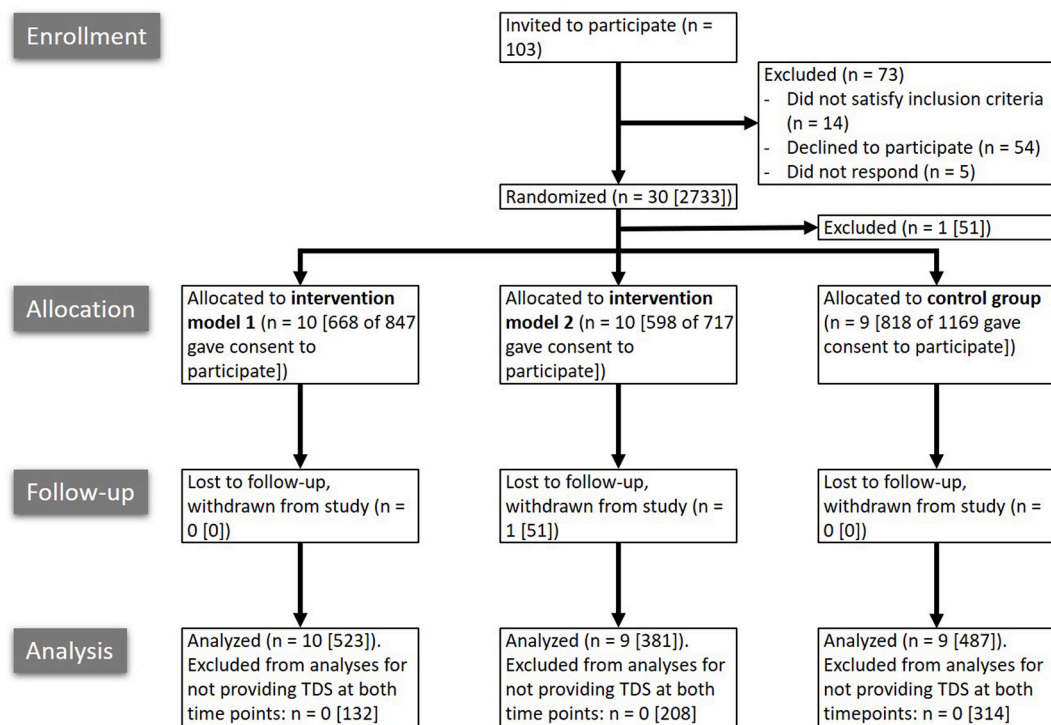


Fig. 1. The flow diagram shows the flow of schools and participants through the School in Motion-study. Numbers represent schools [children]. TDS = Total difficulties score.

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

2.3.1. 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 to 10 and TDS ranges from 0 to 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).

2.3.2. 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.

2.3.3. 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 min 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 s. Overall PA is expressed as average counts per minute (counts \cdot 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 >4 km/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).

2.3.4. 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 pre-determined 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.

2.5. 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).

2.6. 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

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).

Demographics	M1		M2		Control							
	n	%	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								5		
Upper secondary school		27								28		
University < 4 years		43								42		
University > 4 years		24								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.

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.

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 min, 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.

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**.

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)

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

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 = .252)

complete case results.

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%–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, Smith, 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-min 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

4.2.1. 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, Richards, et al., 2016a), 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).

4.2.2. 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, Richards, et al., 2016a): 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 & Dijkstra, 1996) and is inversely associated with hyperactivity (Mastoras, Saklofske, Schwan, & 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.

4.2.3. 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 (Breistol, 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.

Declaration of competing interest

The authors declare that they have no competing interests.

Acknowledgments

Thanks to all schools, principals, teachers, students and parents, who agreed to participate. The research was funded by grants from The Norwegian Directorate for Education and Training.

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 and 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 theory and the concept of embodiment. *Advances in Child Development and Behavior*, 45, 261–286. <https://doi.org/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. <https://doi.org/10.1093/jpepsy/jsq107>.
- Amrhein, V., Greenland, S., & McShane, B. (2019). Scientists rise up against statistical significance. *Nature*, 567, 305–307. <https://doi.org/10.1038/d41586-019-00857-9>.
- Åvitsland, A., Leibinger, E., Solberg, R., Lerum, Ø., Haugen, T., Kolle, E., et al. (2020a). *The association between physical fitness and mental health in Norwegian adolescents*. 20. BMC Public Health. <https://doi.org/10.1186/s12889-020-08936-7>.
- Åvitsland, A., Ohna, S. E., Dyrstad, S. M., Tjomsland, H. E., Lerum, Ø., & Leibinger, E. (2020b). The process evaluation of a school-based physical activity intervention: Influencing factors and potential consequences of implementation (2nd ed., 120). *Health Education*. <https://doi.org/10.1108/HE-01-2020-0004>.
- 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 and Behavioral Pediatrics*, 30(5), 376–384. <https://doi.org/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. <https://doi.org/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/Publikasjoner/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. <https://doi.org/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. *International Migration Review*, 36(2), 389–415. <https://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. <https://doi.org/10.1136/bmj.314.7080.572>.
- 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>.
- Bonhauer, M., Fernandez, G., Püschel, K., Yañez, F., Montero, J., Thompson, B., et al. (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. <https://doi.org/10.1093/heapro/dah603>.
- Breistol, 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. <https://doi.org/10.21307/sjcap-2017-003>.
- 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. <https://doi.org/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. <https://doi.org/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. <https://doi.org/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. <https://doi.org/10.1136/bmj.a1655>.
- Dalene, K. E., Anderssen, S. A., Andersen, L. B., Steene-Johannessen, J., Ekelund, U., Hansen, B. H., et al. (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. <https://doi.org/10.1111/sms.12876>.
- 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>.
- 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. <https://doi.org/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](https://doi.org/10.1016/0193-3973(91)90029-4).
- Edbom, T., Lichtenstein, P., Granlund, M., & Larsson, J. (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. <https://doi.org/10.1111/j.1651-2227.2006.tb02311.x>.
- Erola, J., Jalonén, 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. <https://doi.org/10.1017/s0033291714002888>.
- Fröjd, S. A., Nissinen, E. S., Pelkonen, M. U. I., Marttunen, M. J., Koivisto, A.-M., & Keltiala-Heino, R. (2008). Depression and school performance in middle adolescent boys and girls. *Journal of Adolescence*, 31(4), 485–498. <https://doi.org/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, R. (1997). The strengths and difficulties questionnaire: A research note. *Journal of Child Psychology and Psychiatry*, 38(5), 581–586. <https://doi.org/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>.
- 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>.
- Greenland, S., Senn, S. J., Rothman, K. J., Carlin, J. B., Poole, C., Goodman, S. N., et al. (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., et al. (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. <https://doi.org/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. *Current Opinion in Psychiatry*, 20(4), 359–364. <https://doi.org/10.1097/YCO.0b013e32816ebc8c>.
- Killip, S., Mahfoud, Z., & Pearce, K. (2004). What is an intraclass 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 in a population-based sample of Norwegian 9- and 15-year-olds. *Scandinavian Journal of Medicine & Science in Sports*, 20(1), e41–e47. <https://doi.org/10.1111/j.1600-0838.2009.00892.x>.
- Kolle, E., Steene-Johannessen, J., Säfvenbom, R., Grydeland, M., Tjømsland, 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/00554e6be9104daeb387287132cef1e0sluttrapport-scim.pdf>.
- Kolle, E., Solberg, R. B., Säfvenbom, R., Dyrstad Sindre, 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*.
- 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. (2016a). Physical activity for cognitive and mental health in youth: A systematic review of mechanisms. *Pediatrics*, 138(3), Article e20161642. <https://doi.org/10.1542/peds.2016-1642>.
- Lubans, D. R., Smith, J. J., Morgan, P. J., Beauchamp, M. R., Miller, A., Lonsdale, C., et al. (2016b). 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., Schween, V., & Climie, E. (2015). Social support in children with ADHD: An exploration of resilience. *Journal of Attention Disorders*, 22. <https://doi.org/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. <https://doi.org/10.1111/sjop.12021>.
- Moore, G. F., Evans, R. E., Hawkins, J., Littlecott, H., Melendez-Torres, G. J., Bonell, C., et al. (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. <https://doi.org/10.1177/1356389018803219>.
- Moran, M. D. (2003). Arguments for rejecting the sequential Bonferroni in ecological studies. *Ecology*, 84(2), 403–405. <https://doi.org/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](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. *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., & Roysamb, 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. <https://doi.org/10.1111/j.1467-9450.2004.00388.x>.
- Pels, F., & Kleiner, J. (2016). Loneliness and physical activity: A systematic review. *International Review of Sport and Exercise Psychology*, 9, 1–30. <https://doi.org/10.1080/1750984X.2016.1177849>.
- Perneger, T. V. (1998). What's wrong with Bonferroni adjustments. *BMJ*, 316(7139), 1236–1238. <https://doi.org/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. <https://doi.org/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. <https://doi.org/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., & Sjøgaard, 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. *The 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. <https://doi.org/10.1007/s11336-008-9101-0>.
- Singh, A. S., Saliassi, E., van den Berg, V., Uijtdeuwilgen, 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. <https://doi.org/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 and 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. *BMJ*, 338, b2393. <https://doi.org/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. <https://doi.org/10.1177/0962280216683570>.
- Sun, X., Briel, M., Busse, J. W., You, J. J., Akl, E. A., Mejza, F., et al. (2012). Credibility of claims of subgroup effects in randomised controlled trials: Systematic review. *BMJ*, 344, e1553. <https://doi.org/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. *Journal of Child Psychology and Psychiatry*, 49(12), 1304–1312. <https://doi.org/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. <https://doi.org/10.1111/j.1467-9450.2004.00374.x>.
- Wang, R., & Ware, J. H. (2013). Detecting moderator effects using subgroup Analyses. *Prevention Science*, 14(2), 111–120. <https://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., Fornea, 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. <https://doi.org/10.1111/1467-9507.00195>.
- Youthmind. (2018). Information for researchers and professionals about the strengths & difficulties questionnaires. Retrieved from <http://www.sdqinfo.com/>.