Need-Supportive Professional Development in Elementary School Physical Education: Effects of a Cluster-Randomized Control Trial on Teachers' Motivating Style and Student Physical Activity

Géraldine Escriva-Boulley[1], Damien Tessier[1], Nikos Ntoumanis[2], & Philippe Sarrazin[1]

1. Univ. Grenoble Alpes, SENS, F-38000 Grenoble, France

2. Curtin University, School of Psychology and Speech Pathology, Perth, Australia

Corresponding author:

Dr. Damien Tessier, Laboratoire Sport et Environnement Social (EA 3742), University Grenoble-Alpes, BP 53 – 38041 Grenoble cedex 9, France. E-mail: damien.tessier@univ-grenoble-alpes.fr

Abstract

This cluster-randomized controlled trial investigated the efficacy of a teacher professional development (TPD) program, grounded on self-determination theory, to increase elementary school teachers' need-supportive motivating style and consequently their students' physical activity (PA) during PE lessons. Participants were 15 elementary school teachers and their 293 students. Teachers in the treatment condition received a sports-related notebook and attended four 3-hours workshops over one school year; teachers in the control condition received the notebook only. Students' PA and teachers' motivating style were assessed on four occasions via accelerometers and observations, respectively. Results showed that teachers in the treatment condition increased support of their students' psychological needs for the majority of the school year, but there was a slight decrease in the fourth wave of measurement. Students in the treatment condition increased their time spent in moderate to vigorous PA (MVPA), independently of the sport taught, while their counterparts from the control condition decreased their MVPA. This is the first study to provide elementary school teachers with a PE TPD program grounded in SDT, and demonstrate the potential of such a program to improve teachers' motivating style and student MVPA in PE.

Keywords: self-determination theory; teacher training; motivation; cluster-randomized control trial; intervention

There is a large body of evidence demonstrating the beneficial effects of physical activity (PA) on physical and psychological health in youth (Janssen & LeBlanc, 2010). In order to reap these benefits, it is recommended that young people participate in at least 60 minutes of moderate to vigorous intensity of PA (MVPA) every day (World Health Organisation, 2010). However, only 19% of young people undertake PA at a level that meets these PA guidelines (World Health Organisation, 2010). To address this problem, the school setting has been recommended as a key environmental context for PA promotion (Pate, Davis, Robinson, Stone, McKenzie, & Young, 2006). In particular Physical Education (PE), being a compulsory subject that includes all members of an age cohort, is an important setting to help youth to engage in PA at levels that contribute toward meeting current PA recommendations (Bassett et al., 2013). To this end, it is important that schools provide regularly scheduled PE classes and that students are appropriately motivated to fully engaged in them. Unfortunately, children spend on average only 32.6% of PE lesson time engaging in MVPA, as assessed by accelerometers (Hollis et al., 2016), well below the recommended 50% target (Center for Disease Control and Prevention, 2010). Thus, interventions to increase the percentage of time elementary school students spend on MVPA during PE classes are needed (Hollis et al., 2016).

In a systematic review Lonsdale and colleagues (2013) found that engagement in teacher professional development (TPD) programs is an effective strategy to increase student PA in PE. However, less than half of the intervention studies included in that review utilized a theoretical framework to explain student behavior. Hence, Lonsdale et al. stressed the necessity for these studies to be informed by motivational theories that have proved useful in explaining student behavior in PE, such as Self-Determination Theory (SDT; Deci & Ryan, 2002). In the present intervention study we tested whether providing elementary school teachers with a TPD program grounded in SDT, which aimed to facilitate elementary school

teachers' adoption of a more motivating style during PE lessons, would increase their students' PA.

Fostering Student MVPA in PE: The Role of Teachers' Need-Supportive Motivating Style

Over the last two decades, SDT has been shown to be a heuristic theoretical framework for the investigation of motivation in PE classes, explaining how teachers can improve their students' MVPA in lessons as well as their motivation (see Owen, Smith, Lubans, Ng, & Lonsdale, 2014; Van den Berghe, Vansteenkiste, Cardon, Kirk, & Haerens, 2014 for overviews). In the SDT framework students are purported to have three psychological needs — autonomy (e.g., the need to experience a sense of choice and freedom to engage in an activity), competence (e.g., the need to feel able to effectively realize challenging tasks set out by the teacher), and relatedness (e.g., the need to develop meaningful relationships with their teacher or classmates). The satisfaction of these needs facilitates student engagement in PE (Wilson, et al., 2012).

A critical tenet of SDT is that teachers have the capacity to support or thwart students' basic psychological needs, depending on their motivating style during instruction (Reeve, 2009). Teachers' motivating style refers to "the interpersonal sentiment and behavior a teacher uses to motivate his or her students to engage in learning activities" (Reeve et al., 2014, p. 94). A need-supportive motivating style is characterized by the provision of autonomy support, structure, and involvement, intended to nurture students' psychological needs for autonomy, competence and relatedness, respectively (Reeve, 2009; Reeve, Jang, Carrell, Jeon, & Barch, 2004; Skinner & Edge, 2002). By contrast, a need-thwarting motivating style is characterized by controlling, chaotic and hostile teaching behaviors, that are assumed to undermine students' psychological needs for autonomy, competence and

relatedness, respectively. Some studies showed that these dimensions of need support and need thwarting are likely to be inversely related, but are not necessarily bipolar (Bartholomew, Ntoumanis, Ryan, & Thøgersen-Ntoumani, 2010; Smith et al., 2015). For instance, a teacher could be close to his/her students and at the same time be chaotic by giving unclear objectives.

In more detail, an *autonomy supportive* teacher facilitates student autonomy by nurturing the students' inner motivational resources, providing them with explanatory rationales, relying on non-controlling and informational language, displaying patience to allow time for self-paced learning, and acknowledging and accepting expressions of negative affect (Reeve, 2009). In contrast, a *controlling* teacher relies on external sources of student motivation, such as directives, outcome-contingent incentives, pressure, or threats of punishment, to get the students to behave in teacher-desired ways. *Structure* "refers to the amount and clarity of information that teachers provide to students regarding what to do and how to do it so as to develop desired skills and to achieve valued outcomes" (Reeve & Cheon, 2014, p. 298). Teachers provide structure by clearly communicating guidelines and expectations to initiate a learning activity (Jang, Reeve, & Deci, 2010; Sierens, Vansteenkiste, Goossens, Soenens, & Dochy, 2009), by offering sufficient guidance during lessons, by providing step-by-step directions following the pace of the learners (e.g., Jang et al., 2010), and by giving positive and constructive feedback (e.g., Koka & Hein, 2005; Mouratidis, Vansteenkiste, Lens, & Sideridis, 2008) to help students build on their skills and sense of competence. In contrast, when a teacher is chaotic, the students' need for competence cannot be satisfied because the students do not know if they are performing skillfully, or because their teacher provides unclear instructions, vague goals, and delivers no informational feedback. Finally, an *involved* teacher invests a considerable amount of time, energy, and resources in their students, and offers affection, unconditional regard, warmth, care, and

nurturance, whereas, a *hostile* teacher is neglectful or even aggressive in his/her interactions with students (Skinner & Edge, 2002).

A need-supportive motivating style has been found to be related to several positive outcomes in PE classes (see Ntoumanis, 2012, Van den Berghe et al., 2014, for reviews), including higher engagement (e.g., Van den Berghe, Cardon, Tallir, Kirk, & Haerens, 2015), health-related well-being (Standage, Gillison, Ntoumanis & Treasure, 2012) and MVPA (e.g., Perlman, 2013). In contrast, a need-thwarting motivating style has been related to student disengagement (Van den Berghe et al., 2015), fear of failure and challenge avoidance (Bartholomew et al., in press). While an abundant amount literature supports the benefits for students of a need-supportive motivating style, there are studies showing that this style is not frequently in operation in PE lessons (Haerens et al., 2013). In fact, PE teachers tend to mainly use a controlling/thwarting motivating style (Sarrazin, Tessier, Trouilloud & Chanal, 2006). Therefore, it is important to examine whether teachers can be trained to adopt a need-supportive motivating style and to avoid a need-thwarting motivating style, and whether such changes in teacher behaviors are associated with increases in student MVPA.

Intervention Studies Using an SDT Framework

The critical question posed by studies applying SDT in PE is not only whether teachers can adopt a need-supportive motivating style, but also whether their students can benefit from it in terms of their MVPA during PE lessons. A meta-analysis by Su and Reeve (2011) of SDT-based studies in different life domains, including schools, showed that autonomy-supportive intervention programs are effective in changing supervisors' motivating style with a mean weighted effect size of 0.63. The effect size showed some variability as a function of the level of experience, and was higher for inexperienced trainees than experienced ones. In addition, Su and Reeve identified six design features that were common

to the most effective intervention programs. Specifically, these programs (1) offered a workshop that featured all aspects (rather than only one or a few) of an autonomy-supportive style (see Reeve, 2009), (2) delivered training in multiple (rather than a single) sessions, (3) included a group discussion component where teachers could express their reservations and also exchange ideas regarding instructional strategies, (4) offered teachers ongoing support throughout the intervention implementation period, (5) emphasized not only content (what to do) but also skill-based training (how to do it), and (6) addressed teachers' beliefs about motivating style and effective ways of motivating others that might conflict with the training content (see also Ntoumanis, Quested, Reeve, & Cheon, in press).

In the PE context, only a few intervention studies to date have been carried out (Aelterman, Vansteenkiste, Van den Berghe, De Meyer, & Haerens, 2014; Chatzisarantis & Hagger, 2009; Cheon & Reeve, 2013; 2015; Cheon, Reeve, & Moon, 2012; Lonsdale et al., 2016; Perlman, 2015; Tessier, Sarrazin, & Ntoumanis, 2008, 2010). On the whole, these studies examined the effects of a need-supportive motivating style training program on teacher behavior, students' basic need satisfaction and motivation in PE, and a variety of student-related cognitive, affective, and behavioral outcomes. With the exception of Tessier et al. (2010), these intervention studies compared teachers in an SDT-informed treatment condition with those in a control condition (i.e., standard teaching practice). Two general findings have emerged from such studies: trained teachers can learn how to become more autonomy supportive (e.g., Chatzisarantis & Hagger, 2009; Cheon & Reeve, 2015; Cheon et al., 2012; Perlman, 2015; Tessier et al., 2008) or need-supportive (e.g., Aelterman et al., 2014; Tessier et al., 2010) and, secondly, by doing so, trained teachers provide their students with better lesson experience that leads to gains in terms of student engagement (Cheon & Reeve, 2015; Cheon et al., 2012) and intention to exercise during leisure time (Chatzisarantis & Hagger, 2009; Cheon et al., 2012).

However, this body of knowledge is not without its limitations. First, intervention studies focusing on the impact of teachers' motivating style on their students' MVPA are lacking. Previous intervention studies have demonstrated increased student engagement at the class level (i.e., by measuring collective engagement of the whole class; e.g., Cheon et al., 2012; Tessier et al., 2010) using self-report measures of behavioral, emotional, cognitive, and agentic engagement. However, student behavioral engagement, in terms of accelerometry-assessed MVPA, was not reported in these studies. Given the health benefits derived from MVPA it is important to know if a SDT-informed TPD program is effective in increasing student MVPA in PE. Second, most past intervention studies (with the exception of Lonsdale et al., 2016; Tessier et al., 2010) adopted a narrow conceptualization of motivating style that focused only on autonomy support or autonomy support and structure (Aelterman et al., 2014), rather than on all three aspects of need-support (i.e., autonomy support, structure, and involvement). Given that these are complementary rather than independent dimensions of a teacher's style (e.g., Jang, et al., 2010), it is important to include all three dimensions in a TPD program. A final limitation is that almost all previous studies (with the exception of Chatzisarantis & Hagger, 2009) have not addressed the possibility that participants in a treatment condition might benefit from a Hawthorne effect – i.e., the tendency to work harder merely because of the additional attention paid to them by the researcher (McCambridge, Witton, & Elbourne, 2014). To prevent such an effect, teachers from the control condition should also receive attention by being offered a TPD program.

The Present Study

The purpose of this intervention study was to examine whether an 8-month long SDT-based TPD program, aimed at training elementary school teachers to adopt a more need-supportive motivating style during PE lessons, would have a positive impact on teachers' motivating style and on their students' MVPA. This study built upon and expanded in six

ways, conceptually and methodologically, on previous studies carried out in the general education context, and in PE in particular. From a conceptual perspective, rather than focusing on autonomy-support only, this study targeted all three need-supportive (i.e., autonomy support, structure, involvement) and three need-thwarting (i.e., control, chaos, hostility) motivating styles. Secondly, the TPD program took into account all six design features that maximized the effectiveness of intervention programs in the meta-analytic review by Su and Reeve (2011). From a methodological perspective, given that children under 10 years of age have problems recalling their PA retrospectively (Corder, Ekelund, Steele, Wareham, & Brage, 2008), accelerometers were used in this study to obtain non-biased estimates of the percentage of PE lesson time spent on MVPA. Also, to prevent a potential Hawthorne effect (McCambridge et al., 2014), teachers in the control condition were also provided with TPD. Similar to past TPD programs in which teachers had access to learning and teaching resources and PE lesson plans (e.g., Morris et al., 2013), in the present study teachers in the control condition were given notebooks presenting a variety of learning tasks. In addition, this study was carried out with elementary school teachers and elementary-aged children. In contrast to secondary PE teachers who are PE specialists, in the French education system elementary school teachers have several compulsory subjects to teach (e.g., mathematics, history and geography, French, foreign languages), and thus are not specialist in the subject of PE[1]. As a result, the effectiveness of TPD programs among non-specialist teachers is not known. Finally, the present study used observations to assess teachers' motivating style which allows for the identification of concrete, real-life examples of how teacher need support manifests in the classroom (Haerens et al., 2013). Additionally, one strength of this method is to provide treatment fidelity information via the changes in the observation scores (Cheon & Reeve, 2015).

It was hypothesized that, elementary school teachers who received the TPD program would increase their need-supportive motivating style (i.e., autonomy support, structure and involvement) across the school year, and decrease their need-thwarting motivating style (i.e., controlling, chaotic and hostile teaching behaviors). In addition, given that the teachers from the control condition received notebooks with a number of examples of learning situations and detailed instructions and material, it was predicted that these teachers, similar to the teachers in the treatment condition, would improve their ability to support the students' need for competence (but not for autonomy or relatedness). Indeed, based on the aforementioned definition of 'structure', it was expected that the notebook would help teachers in both conditions to provide better structure in their PE lesson by being more effective in giving their students information regarding activity planning (what activities, how and when to offer those). Finally, it was expected that the students of the teachers in the treatment condition would benefit from the intervention and display higher percentage of PE lesson spent in MVPA, compared to their counterparts whose teachers were allocated to the control condition.

Method

Participants and Procedure

To reach 80% power to detect an effect of d=0.85 on students' MVPA (Lonsdale et al., 2017), 46 students would have been needed in a non-clustered trial (two-tailed probability level of 0.05). We adjusted our calculations for class level clustering using Campbell, Elbourne and Altman's (2004) formula $1 + (m-1)\rho$, where m is the number of students per class and ρ is the intra-class correlation. With an estimated class size of 22 participating students (mean number of student in primary school in France) and an intra-class correlation of 0.07 (Lonsdale et al., 2017), an adjustment of 2.47 was required: $1 + (22–1)*0.07 = 2.47$. Multiplying by the 46 participants required in non-clustered trial, 113 students would have

been needed to achieve 80 % power. With regard to the number of teachers needed, on the basis of the mean weighted effect size (d=0.63) observed in the Su and Reeve's (2011) meta-analysis, we calculated using the G*Power (Faul, Erdfelder, Lang, & Buchner, 2007) software that 82 teachers would have been needed to reach a power of .80. Two hundred and forty nine French elementary school administrators were contacted by email at the beginning of the school year. The TPD was also included in the pool of TPDs made available to teachers by school authorities. Fifteen full-time certified elementary school teachers (66.6% females, $M_{\text{teaching experience}}$ = 9.5 years; SD= 6.3; range = 2 years - 23 years) and their 293 students (51.8% of girls; M_{age}= 8.31; SD= 1.13; range= 5 years - 11 years) from 13 state elementary schools agreed to participate to the study. Six teachers were from urban schools (i.e., three in the control condition and three in the treatment condition) and nine from rural schools (i.e., five in the control condition and four in the treatment condition). Seven teachers were from low socio-economic status (SES) schools (i.e., three in the control condition and four in the treatment condition) and eight teachers were from high SES schools (i.e., five in the control condition and three in the treatment condition). The teachers taught at all French elementary school levels (i.e., from first to fifth grade).

The study was not prospectively registered, but ethical approval was obtained from the university ethics committee and the local education authority. Further, consent to participate in the study was obtained from the head teachers of the schools, teachers, and parents. An opt-out procedure was used to obtain parents' consent. A consent form was given to the parents via the students. It provided information about the purpose of the study and its protocol. Parents who did not allow their children to participate had to return it, but none did so. Students could also refuse to participate by saying so to their teacher or to the researchers. No student refused to participate. This study took place during one entire school year (see the timeline in Figure 1 in the supplementary material). It was a parallel group, two-condition,

superiority trial. Schools were assigned randomly to either a control or intervention condition (seven classes to the treatment condition and eight classes to the control condition) with a 1:1 allocation ratio. In order to reduce the risk of contamination, the first author assigned all classes from the same school to the same condition (either treatment or control). The CONSORT extension for clustered trials checklist is available as supplementary material. The teachers from the treatment condition received 12 hours of TPD – 3 hours in October and 3 hours a week before each school vacation. Baseline assessments at the beginning of the school year were not possible because of the delay in obtaining study approval from the local education authority. Thus, the study started one month after the beginning of the school year (i.e., beginning of October).

In France, the school year is divided into five periods of approximately eight weeks, separated by two weeks of vacation. In PE, each period or cycle is focused on the teaching of a particular PA or sport. Consequently, teachers have to plan teaching sequences during each period, focused on specific sports. Teachers in both conditions were asked by the research team to teach four sports, namely orienteering, ultimate Frisbee, dance and rugby, at the same periods. Given that the training was intended to illustrate the need-supportive strategies in a particular teaching cycle (i.e., related to the specificities of the sport activity to be taught), four training sessions were delivered at the end of the four first periods, each one being related to the sport the teachers had to teach during the next period (see Figure 1 in the supplementary material). In PE, teaching is structured to allow students to explore several sports and to learn motor and technical skills. Consequently, teachers have to plan teaching sequences during each period, focused on specific sports. The data collection (i.e., the recording of student MVPA and teachers' behavior) was carried out in the middle of each of the four periods. All teachers received a notebook for each sport targeted which contained detailed instructions and material that allowed the teachers to identify the skills had to

develop, which were appropriate for the grade level and the curriculum taught. The notebook also contained a systematic timetable for each sport (e.g., warm-up, technical drills, tactical situation, game play) that standardized the time allocated to low vs. high intensity activities. No SDT-related material was included in the notebooks. Only the teachers in the treatment condition attended the TPD program. The components of the TPD are presented in the supplementary material.

As with most longitudinal research, there was some missing data over the four time points (see the flow diagram in Figure 2 in the supplementary material). All the teachers provided data at least twice, but most of them were assessed three times. The teachers who completed two or three waves of measurement did not differ significantly from the teachers who completed four waves of measurement on demographic variables, Fs (1, 14) < 0.90, ps > .45, and on raters' scores of motivating style at time 1, Fs (1, 14) < 2.53, ps > .10.

Components of the Professional Development

The teachers in the treatment condition were invited to participate in a four × 3-hour TPD program, offered by a researcher specialist in PE teaching and sport psychology. Each of the four TPD sessions was divided in 3 parts. In every TPD session, the first part was dedicated to presenting basic tenets of SDT, the second one consisted of implementing, in the targeted sport, the need supportive strategies proposed during the first part of the session, and the third one consisted of implementing these strategies in practice (i.e., a teacher enacted the situation with another teacher playing the role of the student). In addition to the TPD sessions, teachers in the treatment condition benefited from onsite follow-up and individualized guidance (for more detail, see the supplementary material).

Measures

Teachers' motivating style. To assess the six dimensions of teachers' motivating style, an observational instrument grounded in SDT and Skinner and Edge's (2002) work was

used. This tool comprises six specific dimensions: “Autonomy support” (e.g., providing choices, rationale), “control” (e.g., controlling use of rewards and language), “structure” (e.g., delivering instructions for the learning situation), “chaos” (e.g., giving no feedback), “involvement” (e.g., showing care/concern), and “hostility” (e.g., belittling students). Two coders blinded to the treatment conditions were provided with a list of exemplary behavioral strategies for each dimension in order to facilitate accurate and reliable rating. Coders were asked to rate each dimension using a potency scale ranging from 0 (*not at all*) to 3 (*strong potency*), similar to the observational instrument developed by Smith et al. (2015) in the sport context. The potency refers both to the frequency (i.e., the number of behavioral strategies used for each dimension during the coding interval) and the quality (i.e., the pervasiveness of the environment in terms of it motivational meaning) of each strategy. More details about this distinction are provided by Smith et al. (2015).

The coder training and coding protocol were identical to those developed by Smith et al. (2015). Teachers in both conditions were videotaped with a digital camcorder, and their communications with students were recorded using a small microphone fixed on the collar of their sweatshirt. During the recording of the PE lesson there was no interaction between the researcher and the participant. In French primary schools teachers schedule the PE lesson, as well as other subjects, whenever they want during the day. As a result, in our study there was variability in the duration of PE lessons, from 23 minutes to 76 minutes. For coding purposes, each video of a PE lesson was divided into four equal quarters to allow for a sufficient time period on which to base the potency ratings. The coders then summed the scores of the four equal quarters in order to obtain a score out of 12 for each of the six dimensions of the motivating style (further details are available in the supplementary material).

Only 45 videos of the 60 intended were obtained. Missing data was not due to teacher attrition but due to technical problems (i.e., video sound problems) and scheduling problems

(i.e., visits cancelled because teachers were ill and it was impossible to reschedule). Results of the inter-rater agreement analysis on the 45 videos were acceptable (average weighted k= 0.92; ranged from 0.77 to 1, see Table 1 in the supplementary material).

Student physical activity levels. To assess student PE lesson time spent in MVPA, each student wore a portable accelerometer (SenseWear® pro2 Armband 6.1 BodyMedia, INC., PA, USA) from the beginning to the end of the PE lesson. The SenseWear Armband has been validated with children between 3 to 10 years old (e.g., Andreacci, Dixon, Dube, & McConnell, 2007; Vorwerg Petroff, Kiess, & Blüher, 2013). The sampling frequency was 1-min epochs. Outcome variables extracted from the accelerometer data were minutes in PA at different intensities (e.g., moderate and vigorous). MVPA was calculated as the sum of moderate and vigorous PA. Given that the duration of the PE classes was different from one lesson to another, the raw MVPA score was converted into a percentage of time spent in MVPA during the PE lesson. Additionally, given that the sport taught could affect student PA levels (see Fairclough & Stratton, 2006, for a review), the percentage of time spent in MVPA during the PE lesson was centered for each wave of measurement to control for "type of sport" effect. In other words, at each wave of measurement, the mean percentage of time spent in MVPA in the whole sample (i.e., both the treatment condition and the control condition data) was subtracted from each student MVPA score. As a result, the group-centered scores from one time point to the next (see Figure 3 in the supplementary material) could be quite different from the percentages of MVPA at the corresponding time points.

Data Analysis

Due to the nested nature of the data, multilevel analyses were performed to test our hypotheses, using the SPSS software, version 21 (SPSS Inc., NY). To examine change in teachers' motivating style across time, data were treated as a two-level hierarchical model (i.e., the four waves of measurement at level 1 and teachers at level 2). To test the change in

the percentage of time student spent in MVPA in PE lesson, data were treated as a three-level hierarchical model (i.e., the four waves of measurement at level 1, students at level 2, and teachers at level 3).

Following the strategy suggested by Singer and Willet (2003), several models were tested. In a preliminary step an unconditional model (model 1) was tested – with only an intercept and no explanatory variables – to partition the variance of each dependent variable into within-individual and between-individual components. This model presents the mean for the whole sample across the whole duration of the study. In step 2, the variable 'time' was included in an unconditional linear growth curve model (model 2) as a fixed parameter. This variable was centered on the first measurement (i.e., four waves of measurement, with wave 1 coded as 0, wave 2 coded as 1; wave 3 coded as 2 and wave 4 coded as 3) and represented the linear change in the treatment condition over time. In step 3, predictors were added. To control for the effect of teacher gender and teaching experience, as well as the effects of student gender and age, these variables were added in the models. Then, two conditional models were set up: (1) the variable 'condition' (a dummy variable, where the treatment condition= 0 and the control condition=1) and the interaction term time × condition" were added as predictors (model 3, linear model), and (2) the variable 'time²', and the interaction 'time² × condition' (model 4, quadratic model) were added as predictors. Model 4 was used because some previous developmental studies in secondary schools showed that growth trajectories for teacher and student motivation-related variables (e.g., Barkoukis, Ntoumanis, & Thøgersen-Ntoumani, 2010) were nonlinear over time. Consequently, linear (model 3) and quadratic models (model 4) were compared.

The 'condition' effect tested whether the treatment condition and the control condition differed at baseline (time = 0), the interaction 'time × condition' examined whether the linear rate of change over time differed across treatments, and the interaction 'time² × condition'

tested whether the curvilinear rate of change over time differed across treatments (model 4). To compare models, the -2 log likelihood (i.e., likelihood ratio test/deviance test; Heck, Thomas, & Tabata, 2014), was used, with lower values indicating better model fit (more details about multilevel analyses are available in supplementary material). Because of space restrictions, only model 1 and the model that fitted the data best for each dependent variable (i.e., model 4 for motivating style and model 3 for percentage of MVPA) are presented. For linear models, the effect sizes were calculated using the formula proposed by Feingold (2013): $d=(b_{time\times group} \times \text{duration}) / SD_{\text{raw}}$, where $b_{time\times group}$, is the difference in rate of change between conditions per unit of time, duration is the length of the study, and SD_{raw} is the standard deviation of raw scores based on data from the first wave of measurement. For nonlinear models, effect sizes were calculated for each time measurement by working out the difference between the model-estimated means of the two groups at that time and dividing this difference score by its standard deviation (Feingold, 2013). Then, the average of the effect sizes obtained at each time point was calculated to inform about the general effect size of the intervention.

Results

Preliminary Analyses

The statistical assumptions associated with multilevel models were checked by exploring the residuals in the full conditional models. Results indicated relative normality in the distribution of the residuals and no extreme outliers. Furthermore, plotting the residuals against the predicted scores of the dependent variables showed no major signs of heteroscedasticity. Means, standard deviations, and weighted Kappas at each measurement wave and for each condition are shown in Table 1 in the supplementary material.

At time 1, 200 students were sampled (48% male; *M* age = 8.29), 84 in the treatment condition and 116 in the control condition. At time 2, 252 students (51% male, *M age* =

8.34), were present for the data collection, 117 in the treatment and 135 in the control condition. The time 3 sample included 183 students (50% male, *M age* = 8.33), 111 in the treatment and 72 in the control condition. Finally, at time 4, 148 students (52% male, *M age* = 8.40) were present for the data collection, 52 in the treatment and 96 in the control condition. It was not possible to measure the percentage of MVPA at the fourth wave of measurement because most teachers refused to allow their students to wear an accelerometer whist they were playing rugby. Given that rugby is a physical contact sport, teachers were concerned that the accelerometer device could hurt some students during the game. Thus, the analysis of the student MVPA was performed on the three first waves only.

Teachers' Motivating Style in PE Lesson

Interclass correlation coefficients from unconditional models were all above 5%, indicating that there was a hierarchical structure in the data and that multilevel analysis was appropriate (Bryk & Raudenbush, 1992). In relation to controlling style, chaos and hostility, results showed no difference between the two conditions, and absolute values of effect sizes ranged from 0.20 to 0.36 (Table 1). Regarding autonomy support, structure and involvement, results showed that the model that had the lowest -2 log likelihood value was the quadratic model, indicating that the trajectory of each of these variables was curvilinear. Specifically, the difference in the trajectories between the two conditions increased across time (i.e., the scores were not different at the first wave of measurement, then scores for teachers in the treatment condition became higher than scores for teachers in the control condition in the second and third waves of measurement) and then decrease at the last wave of measurement (Table 2). Absolute values of effect sizes for all need supportive dimensions ranged from 0.57 to 0.73.

With respect to teachers' autonomy support, results showed no significant main effects for teacher gender (b = 1.75, p = .13) and experience (b = -0.07, p = .39), indicating that there

were no differences in autonomy support as a function of these variables. Results also showed no significant main effects for condition (b = -0.34, p = .79) indicating that there was no difference in the first wave of measurement between the two conditions. Results also revealed significant main effects of time and time² (b = 4.37 and -1.40, ps < .001, respectively), indicating that autonomy support increased significantly over time in the treatment condition but this increase became slower over time. More importantly, results showed significant time × condition (b = -4.58, p = .006) and time² × condition (b = 1.53, p = .005) interactions, indicating that the slope of change between conditions was different. Visual inspection of Figure 4a in the supplementary material shows that autonomy support increased for the treatment condition during the first three waves of measurement and decreased at the last wave, while it remained relatively stable over time for the control condition.

Regarding teachers' structure, results showed no significant main effects for teacher gender (b = 0.58, p = .52) and experience (b = -0.11, p = .12), indicating that there were no differences in structure as a function of these variables. Results also showed no significant main effect for condition (b = 0.55, p = .61), indicating that there was no difference in the first time of measurement between the two conditions. There was a significant main effect for time and time² (b = 4.82 and -1.34, ps < .001, respectively), meaning that structure increased for teachers in the treatment condition but the increase decelerated over time. Results showed a marginal time × condition interaction (b = -2.41, p = .09), indicating that the linear increase was slightly greater in the treatment condition than in the control condition. There was no time² × condition interaction (b = 0.60, p = .21), indicating that there was no difference between conditions in the curvature of their trajectory of change (Figure 4b in the supplementary material).

Regarding teachers' involvement, results showed no significant main effects for teacher gender (b = 0.93, p = .46) and experience (b = -0.06, p = .52), indicating that there

were no differences in involvement as a function of these variables. Results also showed no significant main effect for condition (b = 1.11, p = .40), indicating that there was no difference in the first wave of measurement between the two conditions. Results further revealed no significant main effects for time and time² (b = 1.64, p = .09; b = -0.38, p = .21; respectively), meaning that teacher involvement did not change in the treatment condition over time. However, there were significant time × condition (b = -3.43, p = .02) and time² × condition (b = 0.98, p = .03) interactions, indicating that the slope of change for each condition was different. Visual inspection of Figure 4c in the supplementary material shows that involvement decreased for the control condition during the first three waves of measurement and then increased, while it remained steady over time for the treatment condition.

With the exception of involvement (χ^2 = 4.669, ps < 0.10), the quadratic model fitted better compared to the linear model (χ^2 > 13.771, ps < 0.001) for the other two need-supportive dimensions (i.e., autonomy support and structure; see Table 2).

Student Physical Activity

With regard to PE lesson time spent in MVPA the unconditional model (model 1) for the three-level model (i.e., the four waves of measurement at level 1, students at level 2, and teachers at level 3) revealed that only 2% of the total variation of MVPA was at level 2 (Table 3). As the variance attributed to inter-individual student differences was trivial (Bryk & Raudenbush, 1992) a two levels model was tested, with the three waves of measurement at level 1 and teachers at level 2. Model 3 (linear change) fit the data best.

Results for centered data (Table 3) showed no significant main effect for student age (b = -1.46, p = .28), indicating that there was no differences in % of lesson time spent in MVPA as a function of this variable. Results also showed a significant main effect for student gender (b = 7.04, p < .001), implying that boys' % of MVPA was higher than that of girls.

Results also showed no significant main effect for condition (b = 6.29, p =.32), indicating that there was no difference in the first wave of measurement between the two conditions. There was a significant main effect for time (b = 5.42, p < .001), meaning that time spent in MVPA increased in the treatment condition over time. In addition, results revealed a significant time × condition interaction (b = -8.70, p < .001), showing that the slope of change between the two conditions was different. Specifically, time spent in MVPA increased for the treatment condition and decreased for the control condition (see Figure 3 in the supplementary material); the absolute value of the effect size was substantial (d = 1.33).

Discussion

The purpose of this study was to test the effects of an SDT-based TPD program, on elementary teachers' motivating style and their students' PA in PE. Results showed that, compared to teachers in the control condition, teachers who attended the TPD program improved their need-supportive motivating style and their students increased their time spent in MVPA.

Effect of the Intervention on Teachers' Motivating Style

Results partially supported our hypothesis. Effect sizes were moderate to large for all dimensions of need support. As expected, structure increased in both conditions but was slightly higher in the treatment condition (Figure 4b in the supplementary material). Thus, the notebook helped teachers to structure students' learning by giving teachers clear goals and instructions, and by outlining contingent and consistent learning situations. However, there was a marginal difference in favor of the treatment condition (p = .09), which means that the TPD brought some added value. Indeed, by providing practical experience and allowing collective discussion with other teachers, the TPD program probably offered a stronger implementation of the strategies that support the students' need for competence.

As predicted, compared to teachers in the control condition, teachers in the treatment condition increased the use of autonomy support and maintained a higher level of involvement over time, however in the fourth wave of measurement the between-group difference decreased. This decrease could be explained either by a regression to the mean effect or by the sport taught by the teachers. For instance, the decrease at the last wave of measurement for autonomy support may be related to the nature of the sport taught by the teachers. The sport taught in the fourth part of the study was rugby. Thus, risks related to the physical contact between players might have focused teachers more on physical safety concerns than on satisfying their students' psychological needs. The support of students' needs requires resources from the teachers (e.g., time, attention, energy, motivation), but when the teachers' resources are used to cope with pressures or external constraints– i.e., students' safety concerns – then, they cannot be invested in supporting student needs (Pelletier, Séguin-Lévesque, & Legault, 2002; Reeve, 2009). In contrast, other sports offer more opportunities for student input. In dance for example, by helping students create choreographies, teachers can encourage students to express their feelings, make their own choices, and take initiatives.

Concerning involvement, it seems that for the treatment condition teachers this variable showed a relatively stable pattern over the year, while for teachers in the control condition this variable decreased. Thus, the TPD program seems to have had a protective effect permitting teachers to maintain their involvement at a consistent level across the year, and prevent the reductions in this variable observed in the control condition. This reduction is consistent with recent results of two studies carried out in academic subjects (i.e., mathematics and English as a foreign language; Maulana, Opdenakker, Stroet, & Bosker, 2013; Stroet, Opdenakker, & Minnaert, 2015). A multitude of contextual (e.g., students' disruptive behaviors, time pressure; see Taylor, Ntoumanis & Smith, 2009; Reeve, 2009) and

personal (e.g., affective states; see Forgas, 2002) parameters could be invoked to explain this trend. A better understanding of how and why teachers change their involvement over the school year could be an interesting avenue for future study.

On the whole, the results are aligned with past literature (e.g., Cheon et al., 2012; Cheon & Reeve, 2013; Tessier et al., 2008, 2010), which showed that SDT-based TPD programs had a positive impact on PE teachers' autonomy support. In addition, the present study extends results of previous studies by showing that an SDT-based TPD program can be beneficial for all need-supportive dimensions, given that the trajectory of change was curvilinear (increased and then plateaued) for autonomy support and structure, and there was no decrease in involvement. Such changes in the treatment condition also imply fidelity to the intervention material; that is teachers delivered the intervention as intended. Finally, this study is the first to demonstrate that such a SDT-based TPD program is able to improve the motivating style of elementary school teachers who are not PE specialist and attended only a few PE courses during their pre-service training.

In contrast to our hypothesis, even though the scores for the need-thwarting dimensions of motivating style were lower in the treatment than in the control condition (Figure 4 d-f in the supplementary material), these differences were not significant. These results could be due to low ratings for these dimensions at the beginning of the study (a floor effect). Moreover these behaviors could be more difficult to change (Reeve, 2009). Research shows that several contextual factors influence whether and to what extent a person in a position of authority will display a controlling motivating style during instruction (Matosic, Ntoumanis, & Quested, 2016; Pelletier et al., 2002; Reeve, 2009). Such factors need to be taken into account by future research because they may affect how new information about motivating styles – such as the one delivered in TPD programs – is understood, integrated, accepted, or rejected (Reeve et al., 2014).

Another explanation for the lack of differences between conditions in need thwarting behaviors is that our TPD was more focused on need-supportive motivating style than on need-thwarting. For example, for need-thwarting behaviors teachers were provided with a detailed presentation of these behaviors and their consequences on students, but there was no practical demonstration of those behaviors. For need-supportive behaviors teachers were provided with the same amount of information but they also benefited from practical examples advice, and discussion as to how to be need-supportive, and they implemented these behaviors during a role play. It is possible that beliefs about need-thwarting behaviors (e.g., that these behaviors can be used concurrently with need-supportive behaviors, or that they are not always detrimental) may have not been adequately addressed in the workshops. Future studies could improve this TPD program by giving more attention to the reduction of a need-thwarting motivating style, given that this style has negative effects on student motivation (e.g., De Meyer et al., 2014; Bartholomew et al., in press).

Effect of the Intervention on Student Physical Activity

Another aim of the current study was the assessment of the effect of the TPD program on the percentage of PE lesson time in which students were involved in MVPA. Accelerometer data showed that this percentage in our sample was higher (i.e., average 60%, Table 1 in the supplementary material) than the percentages presented in previous studies (see Hollis et al., 2016, for a review). However, Hollis et al. (2016) showed that in five of the 13 studies included in their review the mean percentage of PE lesson spent in MVPA was greater than 50%. One of these five studies used accelerometer (mean MVPA was 71%; Verstraete et al., 2007). Our results also showed that boys' MVPA was higher than that of girls. These results are in line with those obtained in previous studies (e.g., Verloigne, et al., 2012).

In accordance with our hypotheses, multilevel analysis showed that the students in the treatment condition increased the time they spent in MVPA during the three waves of

measurement, whereas for those in the control condition MVPA time was decreased. This finding is all the more interesting given that the magnitude of the effect size was large (d = 1.33). In more concrete terms, compared to students in the control condition, those in the treatment condition spent on average 10% more of the PE lesson time engaging in MVPA at the third wave (see Table 1 and Figure 3 in the supplementary material).

It is noteworthy that the percentage of MVPA increased at the same measurement waves as when scores for autonomy support and structure also increased. This finding is in line with previous evidence which showed that the combination of high autonomy support and high structure results in greater student MVPA (Jang et al., 2010). It also suggests that despite the high baseline percentage of MVPA, elementary student MVPA in PE could be sensitive to changes in their teachers' motivating style. Another interesting result relates to the absence of significant random effects at the student level, after the random effect of the class was taken into account (see Table 3). This suggests there were no between-student differences likely to predict students' MVPA and that the potential explanatory variables could be at the teacher/class level (e.g., session organization, teacher objective motivating style during PE lessons). Overall, these results reinforce the idea that students' MVPA in PE is influenced by teachers' behaviors and that the latter could be trained to adopt a more need-supportive style and potentially to avoid a need-thwarting motivating style.

Limitations and conclusion

The present study showed that a PE TPD program grounded on SDT could lead to several teacher and student benefits, but as with all studies, our study had some limitations. The first limitation was the small sample of teachers. Yet, we made substantial efforts to recruit a larger number of teachers by (1) contacting the administrators of 249 schools (with a potential pool of more than 1000 teachers), and (2) adding our TPD in the pool of the 150

TPDs offered by academic authorities to the primary school teachers. Only 15 teachers chose this TPD focused on PE, whereas many more teachers were enrolled in the other ones focused on others school subjects (e.g., French, Maths, Sciences, History). As a result, the study is underpowered at the teacher level and thus we are limited in making conclusions about the effect, or lack of it, of our TPD program (see Button et al., 2013 for an overview of low statistical power issues). Future research is clearly warranted with larger sample sizes. Moreover, additional studies should be carried out to understand why primary school teachers are not interested a lot in PE TPD. A second limitation of this study was that it was not possible to collect data at baseline (i.e., before the first TPD session). Despite the addition of several control variables, we cannot be sure that the treatment condition and control condition were equivalent before the intervention. This issue could have affected our results if one of the conditions had higher scores than the other one before the intervention. However, this is not likely to have been the case, as no difference between the conditions was found in the first wave of measurement, but the differences emerged later on in the study. Additionally, demographical variables (i.e., gender and teaching experience) were included in the models as covariates. The findings revealed that the conditions are equivalent regarding these covariates and that adding these variables in the regression analyzes does not affect the results. A third limitation was that teachers' motivating styles were only assessed by observation. If the observation of teacher styles has several advantages to test the effectiveness of the training, it could be also interesting to examine the sensitivity of the students to the change in their teacher's motivating styles, by using self-reported questionnaires validated for children. In the same vein, in order to have a better understanding of the mechanisms explaining the link between teachers' motivating style and students' PA behaviors, it would be relevant to assess students' psychological needs as a mediator, using a validated scale for this age range. A fourth limitation was that the teachers in the two experimental groups did not receive equal

attention, but a "pragmatic" intervention was tested by giving the teachers in the control condition a minimal TPD program. A fifth limitation was that for security reasons it was not possible to assess MVPA at the fourth wave of measurement. Consequently, the full effect of the intervention for MVPA was not known and mediation analysis could not be performed. In the same vein, the effects of the intervention on leisure time PA were not examined. Future research could examine whether this 'stand alone' intervention is sufficient to increase out-of class PA or whether it is necessary to build comprehensive school PA programs including PE as an intervention component (CDC guidelines, 2013).

In conclusion, the results of the current study shed new light on SDT-based studies in PE by looking at the effects of a TPD program which was theory-based, had a didactical content, delivered multiple sessions of TPD, included a group discussion, offered ongoing support, and addressed teachers' beliefs about motivation. Results showed that (1) providing primary school teachers only with a notebook improved structure but was not enough to change other dimension of motivating style (i.e., results for the control condition group), (2) teachers' autonomy support and structure had an upward trajectory in the SDT-based TPD program, and (3) the students of teachers who benefited from the SDT-based TPD program showed increased physical investment. Given these findings, we conclude that a) SDT represent a valuable theoretical framework to consider in PE TPD programs in primary schools, and b) to have an impact on all need-supportive dimensions it is important to incorporate all six design features for effective SDT-based TPD programs (Su & Reeve, 2011). By showing the effectiveness of a training combining motivational theory with practical lesson planning, role playing, and individualized guidance, the present study yields practical insights to guide the training of elementary school pre-service or in-service teachers.

References

Aelterman, N., Vansteenkiste, M., Van den Berghe, L., De Meyer, J., & Haerens, L. (2014). Fostering a need-supportive teaching style: intervention effects on physical education teachers' beliefs and teaching behaviors. *Journal of Sport and Exercise Psychology, 36*(6), 595–609. doi:10.1123/jsep.2013-0229

Andreacci, J. L., Dixon, C. B., Dube, J. J., & McConnell, T. R. (2007). Validation of SenseWear Pro 2 armband to assess energy expenditure during treadmill exercise in children 7--10 years of age. *Journal of Exercise Physiology online, 10*(4), 35-42.

Bartholomew, K., Ntoumanis, N., & Thøgersen-Ntoumani, C. (2010). The controlling interpersonal style in a coaching context: development and initial validation of a psychometric scale. *Journal of Sport and Exercise Psychology, 32,* 193-216.

Bartholomew, K., Ntoumanis, N., Mouratidis, A., Katartzi, E., Thøgersen-Ntoumani, C., & Vlachopoulos, S., (in press). Beware of your teaching style: A school-year long investigation of controlling teaching and student motivational experiences. *Learning and Instruction*

Barkoukis, V., Ntoumanis, N., & Thøgersen-Ntoumani, C. (2010). Developmental changes in achievement motivation and affect in physical education: growth trajectories and demographic differences. *Psychology of Sport and Exercise, 11*, 83-90. doi: 10.1016/j.psychsport.2009.04.008

Bassett, D., Fitzhugh, E., Heath, G., Erwin, P., Frederick, G., Wolff, D… Stout, A. (2013). Estimated energy expenditures for school-based policies and active living. *American Journal of Preventive Medicine, 44*(2), 108–113. doi:10.1016/j.amepre.2012.10.017

Bryk, A. S., & Raudenbush, S. W. (1992). *Hierarchical linear models: Applications and data analysis methods*. Newbury Park, CA: Sage.

Button, K. S., Ioannidis, J. P., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S., & Munafò, M. R. (2013). Power failure: why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience, 14*(5), 365-376. doi:10.1038/nrn3475

Campbell, M. K., Elbourne, D. R., & Altman, D. G. (2004). CONSORT statement: extension to cluster randomised trials. *Bmj*, *328*(7441), 702-708.

Center for Disease Control and Prevention, Division of Adolescent and School Health (2010). *Strategies to improve the quality of physical education*. Washington, DC: United States Government.

Centers for Disease Control and Prevention (2013). *Comprehensive School Physical Activity Programs: A Guide for Schools*. Atlanta, GA: U.S. Department of Health and Human Services

Chatzisarantis, N. L., & Hagger, M. S. (2009). Effects of an intervention based on self-determination theory on self-reported leisure-time physical activity participation. *Psychology and Health, 24*(1), 29–48. doi:10.1080/08870440701809533

Cheon, S. H., & Reeve, J. (2013). Do the benefits from autonomy-supportive PE teacher training programs endure?: A one-year follow-up investigation. *Psychology of Sport and Exercise*, *14*(4), 508–518. doi: 10.1016/j.psychsport.2013.02.002

Cheon, S. H., & Reeve, J. (2015). A classroom-based intervention to help teachers decrease students' amotivation. *Contemporary Educational Psychology*, *40*, 99–111. doi:10.1016/j.cedpsych.2014.06.004

Cheon, S. H., Reeve, J. M., & Moon, I. S. (2012). Experimentally based, longitudinally designed, teacher-focused intervention to help physical education teachers be more autonomy supportive toward their students. *Journal of Sport and Exercise Psychology, 34*, 365–396. doi:10.1123/jsep.2013-0231

Corder, K., Ekelund, U., Steele, R., Wareham, N., & Brage, S. (2008). Assessment of physical activity in youth. *Journal of Applied Physiology, 105*, 977–987. doi:10.1152/japplphysiol.00094.2008

Deci, E. L., & Ryan, R. M. (2002). *Handbook of Self-Determination Research*. Rochester, NY: University of Rochester Press.

Fairclough, S.J., & Stratton, G. (2006). A review of physical activity levels during elementary school physical education. *Journal of Teaching in Physical Education, 25*, 239–257.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behavior Research Methods, 39, 175-191.

Feingold, A. (2013). A regression framework for effect size assessments in longitudinal modeling of group differences. Review of General Psychology. 17(1): 111–121. doi:10.1037/a0030048.

Forgas, J. P. (2002). Feeling and doing: Affective influences on interpersonal behavior. *Psychological inquiry*, *13*(1), 1-28.

Haerens, L., Aelterman, N., Van den Berghe, L., De Meyer, J., Soenens, B., & Vansteenkiste, M. (2013). Observing physical education teachers' need-supportive interactions in classroom settings. *Journal of Sport and Exercise Psychology*, *35*(1), 3–17. doi: 10.1123/jsep.35.1.3

Heck, R. H., Thomas, S. L., & Tabata, L. N. (2014). *Multilevel and longitudinal modeling with IBM SPSS*. New York:Routledge.

Hollis, J. L., Williams, A. J., Sutherland, R., Campbell, E., Nathan, N., Wolfenden, L., ... & Wiggers, J. (2016). A systematic review and meta-analysis of moderate-to-vigorous physical activity levels in elementary school physical education lessons. *Preventive Medicine, 86*, 34–54. doi: 10.1016/j.ypmed.2015.11.018

Jang, H., Reeve, J., & Deci, E. L. (2010). Engaging students in learning activities: It is not autonomy support or structure but autonomy support and structure. *Journal of Educational Psychology, 102*(3), 588–600. doi:10.1037/a0019682

Janssen, I., & LeBlanc, A. G. (2010). Systematic review of the health benefits of physical activity and fitness in school-aged children and youth. *International Journal of Behavioral Nutrition and Physical Activity*, *7*, 1–16. doi: 10.1186/1479-5868-7-40

Koka, A., & Hein, V. (2005). The effect of perceived teacher feedback on intrinsic motivation in physical education. *International Journal of Sport Psychology*, *36*(2), 91-106.

Lonsdale, C., Lester, A., Owen, K. B., White, R. L., Peralta, L., Kirwan, M., ... & Kolt, G. S. (2017). An internet-supported school physical activity intervention in low socioeconomic status communities: results from the Activity and Motivation in Physical Education (AMPED) cluster randomised controlled trial. *British Journal of Sports Medicine*, Published Online First: October 9, 2017. doi:10.1136/bjsports-2017-097904

Lonsdale, C., Lester, A., Owen, K. B., White, R. L., Moyes, I., Peralta, L., ... & Kolt, G. S. (2016). An internet-supported physical activity intervention delivered in secondary schools located in low socio-economic status communities: Study protocol for the Activity and Motivation in Physical Education (AMPED) Cluster Randomized Controlled Trial. *BMC Public Health*, 16:17. doi: 10.1186/s12889-015-2583-7

Lonsdale, C., Rosenkranz, R.R., Peralta, L.R., Bennie, A., Fahey, P., Lubans, D.R. (2013). A systematic review and meta-analysis of interventions designed to increase moderate-to-vigorous physical activity in school physical education lessons. *Preventive Medicine*, *56*, 152–161. doi:10.1016/j.ypmed.2012.12.004

Maulana, R., Opdenakker, M. C., Stroet, K., & Bosker, R. (2013). Changes in teachers' involvement versus rejection and links with academic motivation during the first year of secondary education: A multilevel growth curve analysis. *Journal of youth and adolescence*, *42*(9), 1348-1371. doi:10.1007/s10964-013-9921-9

Matosic, D., Ntoumanis, N., & Quested, E. (2016). Antecedents of Need supportive and controlling interpersonal styles from a self-determination theory perspective: A review and implications for sport psychology research. In Raab, M., Wylleman, P., Seiler, R., Elbe, A.M., and Hatzigeorgiadis, A., *New Perspectives on Sport and Exercise Psychology* (pp. 145–180). Elsevier. doi: 10.1016/B978-0-12-803634-1.00007-8

McCambridge, J., Witton, J., & Elbourne, D. R. (2014). Systematic review of the Hawthorne effect: new concepts are needed to study research participation effects. *Journal of Clinical Epidemiology, 67*(3), 267-277. doi:10.1016/j.jclinepi.2013.08.015

Morris, J. G., Gorely, T., Sedgwick, M. J., Nevill, A., & Nevill, M. E. (2013). Effect of the Great Activity Programme on healthy lifestyle behaviours in 7–11 year olds. *Journal of Sports Sciences, 31*(12), 1280-1293. doi: 10.1080/02640414.2013.781665

Mouratidis, A., Vansteenkiste, M., Lens, W., & Sideridis, G. (2008). The motivating role of positive feedback in sport and physical education: Evidence for a motivational model. *Journal of Sport and Exercise Psychology*, *30*(2), 240-268.

Ntoumanis, N. (2012). A self-determination theory perspective on motivation in sport and physical education: Current trends and possible future research directions. In G.C. Roberts and D. C. Treasure (Eds). *Motivation in sport and exercise* (vol.3, pp. 91-128). Champaign, IL: Human Kinetics.

Ntoumanis, N., Quested, E., Reeve, J., Cheon, S.H. (in press). Need supportive communication: Implications for motivation in sport, exercise, and physical activity. In B. Jackson, J.A. Dimmock, & J. Compton (Eds.), *Persuasion and communication in sport, exercise, and physical activity*. Abingdon, UK: Routledge.

Owen, K. B., Smith, J., Lubans, D. R., Ng, J. Y., & Lonsdale, C. (2014). Self-determined motivation and physical activity in children and adolescents: A systematic review and meta-analysis. *Preventive medicine*, *67*, 270-279. doi.org/10.1016/j.ypmed.2014.07.033

Pate, R. R., Davis, M. G., Robinson, T. N., Stone, E. J., McKenzie, T. L., & Young, J. C. (2006). Promoting physical activity in children and youth. *Circulation, 114*(11), 1214-1224.

Pelletier, L. G.; Séguin-Lévesque, C.; Legault, L. (2002). Pressure from above and pressure from below as determinants of teachers' motivation and teaching behaviors. *Journal of Educational Psychology*, *94*, 186–196. doi: 10.1037/0022-0663.94.1.186

Perlman, D.J., 2013. The influence of the social context on students' in-class physical activity. *Journal of Teaching in Physical Education*, *34*, 46–60.

Perlman, D. (2015). Assisting preservice teachers toward more motivationally supportive instruction. *Journal of Teaching in Physical Education, 34*, 119–130. doi: 10.1123/jtpe.2013-0208

Reeve, J. (2009). Why teachers adopt a controlling motivating style toward students and how they can become more autonomy supportive. *Educational Psychologist*, *44*, 159–178.

Reeve, J., & Cheon, H. S. (2014). An intervention-based program of research on teachers' motivating styles. In S. Karabenick & T. Urdan's (Eds.), *Advances in motivation and achievement* (Vol. 18, pp. 297–343). Bingley, UK: Emerald Group Publishing. doi:10.1108/S0749-742320140000018008.

Reeve, J., Jang, H., Carrell, D., Jeon, S., & Barch, J. (2004). Enhancing students' engagement by increasing teachers' autonomy support. *Motivation and Emotion, 28*, 147–169. doi:10.1023/B:MOEM.0000032312.95499.6f

Reeve, J., Vansteenkiste, M., Assor, A., Ahmad, I., Cheon, S. H., Jang, H., … Wang, C. K. J. (2014). The beliefs that underlie autonomy-supportive and controlling teaching: A multinational investigation. *Motivation and Emotion, 38*, 93–110. doi:10.1007/s11031-013-9367-0

Sarrazin, P. G., Tessier, D. P., Pelletier, L. G., Trouilloud, D. O., & Chanal, J. P. (2006). The effects of teachers' expectations about students' motivation on teachers' autonomy- supportive and controlling behaviors. *International Journal of Sport and Exercise Psychology*, *4*(3), 283–301. doi: 10.1080/1612197X.2006.9671799

Sierens, E., Vansteenkiste, M., Goossens, L., Soenens, B., & Dochy, F. (2009). The synergistic relationship of perceived autonomy support and structure in the prediction of self- regulated learning. *British Journal of Educational Psychology*, *79*(1), 57-68. doi: 10.1348/000709908X304398

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. New York, NY:Oxford University press.

Skinner, E. & Edge, K. (2002). Parenting, motivation, and the development of children's coping. In L. Crockett (Ed.), *Agency, Motivation, and The Life Course* (vol.48, pp. 77–145). Lincoln, NE: University of Nebraska Press.

Slingerland, M., & Borghouts, L. (2011). Direct and indirect influence of physical education-based interventions on physical activity: A review. *Journal of Physical Activity and Health*, *8*(6), 866-878.

Smith, N., Tessier, D., Tzioumakis, Y., Quested, E., Appleton, P., Sarrazin, P., … Duda, J. L. (2015). Development and validation of the multidimensional motivational climate observation system. *Journal of Sport and Exercise Psychology, 37*(1), 4–22. doi:10.1123/jsep.2014-0059

Standage, M., Gillison, F., Ntoumanis, N., & Treasure, D. (2012). Predicting students' physical activity and health-related well-being: A prospective cross-domain investigation of motivation across school physical education and exercise settings. *Journal of Sport and Exercise Psychology, 34,* 37–60.

Stroet, K., Opdenakker, M. C., & Minnaert, A. (2015). What motivates early adolescents for school? A longitudinal analysis of associations between observed teaching and motivation. *Contemporary Educational Psychology, 42,* 129-140. doi: 10.1016/j.cedpsych.2015.06.002

Su, Y.-L., & Reeve, J. (2011). A meta-analysis of the effectiveness of intervention programs designed to support autonomy. *Educational Psychology Review, 23*, 159–188. doi: 10.1007/s10648-010-9142-7

Taylor. I.,M., & Ntoumanis, N., & Smith, B. (2009). The social context as a determinant of teacher motivational strategies in physical education. *Psychology of Sport and Exercise, 10,* 235-243.

Tessier, D., Sarrazin, P., & Ntoumanis, N. (2008). The effects of an experimental program to support students' autonomy on the overt behaviours of physical education teachers. *European Journal of Psychology of Education, 23,* 239–253. doi: 10.1007/BF03172998

Tessier, D., Sarrazin, P., & Ntoumanis, N. (2010). The effect of an intervention to improve newly qualified teachers' interpersonal style, students' motivation and psychological need satisfaction in sport-based physical education. *Contemporary Educational Psychology, 35,* 242–253. doi : 10.1016/j.cedpsych.2010.05.005

Van den Berghe, L., Cardon, G., Tallir, I., Kirk, D., & Haerens, L. (2015). Dynamics of need-supportive and need-thwarting teaching behavior: the bidirectional relationship with student engagement and disengagement in the beginning of a lesson. *Physical Education and Sport Pedagogy,* 1–18. doi: 10.1080/17408989.2015.1115008

Van den Berghe, L., Vansteenkiste, M., Cardon, G., Kirk, D., & Haerens, L. (2014). Research on self-determination in physical education: key findings and proposals for future research. *Physical Education and Sport Pedagogy, 19,* 97-121. doi: 10.1080/17408989.2012.732563

Verloigne, M., Van Lippevelde, W., Maes, L., Yıldırım, M., Chinapaw, M., Manios, Y., ... & De Bourdeaudhuij, I. (2012). Levels of physical activity and sedentary time among 10-to 12-year-old boys and girls across 5 European countries using accelerometers: an observational study within the ENERGY-project. *International Journal of Behavioral Nutrition and Physical Activity, 9*(1), 34-42. doi: 10.1186/1479-5868-9-34

Verstraete, S. J., Cardon, G. M., De Clercq, D. L., & De Bourdeaudhuij, I. M. (2007). A comprehensive physical activity promotion programme at elementary school: the effects on physical activity, physical fitness and psychosocial correlates of physical activity. *Public health nutrition, 10*(5), 477-484.

Vorwerg, Y., Petroff, D., Kiess, W., & Blüher, S. (2013). Physical activity in 3–6 year old children measured by SenseWear Pro®: direct accelerometry in the course of the week and relation to

weight status, media consumption, and socioeconomic factors. *PLoS One*, *8*(4), e60619. doi: 10.1371/journal.pone.0060619

Wilson, A. J., Liu, Y., Keith, S. E., Wilson, A. H., Kermer, L. E., Zumbo, B. D., & Beauchamp, M. R. (2012). Transformational teaching and child psychological needs satisfaction, motivation, and engagement in elementary school physical education. *Sport, Exercise, and Performance Psychology*, *1*(4), 215-230. doi: 10.1037/a0028635

World Health Organisation. (2010). *Global recommendations on physical activity for health.*

Table 1

Results of the Multilevel Models for Control, Chaos, and Hostility

		Control	Chaos	Hostility
	Variables	*b (SE)*	*b (SE)*	*b (SE)*
Fixed Effect	Intercept	3.42 (1.13)*	1.07 (1.21)	2.82 (1.08)*
	Sex	0.84 (1.02)	-1.48 (1.23)	-0.77 (0.84)
	Experience	-0.13 (0.08)	0.13 (0.10)	-0.07 (0.07)
	Condition	1.34 (1.27)	0.85 (1.32)	1.88 (1.28)
	Time	0.34 (1.21)	-0.58 (0.98)	-0.60 (1.44)
	Time²	-0.11 (0.38)	0.28 (0.30)	0.29 (0.46)
	Time x Condition	2.17 (1.72)	0.23 (1.40)	3.06 (2.07)
	Time² x Condition	-0.76 (0.56)	-0.09 (0.45)	-1.25 (0.67)
Random Effect (Model 1)	Level 1	3.81 (1.03)***	2.13 (0.57)***	5.67 (1.52)***
	Level 2	2.61 (1.54)^{t}	4.99 (2.13)*	1.27 (1.35)
Random Effect (Model 4)	Level 1	3.18 (0.85)***	2.08 (0.57)***	4.48 (1.24)***
	Level 2	1.83 (1.14)^{t}	3.60 (1.71)*	0.31 (0.83)
Test of significance				
Reference model		190.350	181.044	194.891
Δ-2logV		186.138	179.939	191.072
χ^2 (df)		4.212 (2)	1.105 (2)	3.819 (2)
Effect size |d|		0.36	0.20	0.30

Notes. Model 4: Treatment Condition = 0 and Control Condition = 1. Sex: male=1

* ^{t} < .10, **p* < .05, ***p* < .01, ****p* < .001

Table 2

Results of the Multilevel Models for Autonomy Support, Structure, and Involvement

		Autonomy support	Structure	Involvement
	Variables	*b (SE)*	*b (SE)*	*b (SE)*
Fixed Effect	Intercept	4.20 (1.13)***	5.61 (0.96)***	5.20 (1.20)***
	Sex	1.75 (1.08)	0.58 (0.88)	0.93 (1.22)
	Experience	-0.07 (0.08)	-0.11 (0.07)	-0.06 (0.10)
	Condition	-0.34(1.25)	0.55 (1.08)	1.11 (1.30)
	Time	4.37 (1.07)***	4.83 (1.00)***	1.64 (0.95)
	Time²	-1.40 (0.34)***	-1.34 (0.31)***	-0.38 (0.30)
	Time x Condition	-4.58 (1.53)**	-2.42 (1.43)	-3.43 (1.35)*
	Time² x Condition	1.53 (0.50)**	0.60 (0.46)	0.98 (0.44)*
Random Effect (Model 1)	Level 1	4.02 (1.07)***	4.49 (1.19)***	2.49 (0.66)***
	Level 2	3.96 (2.02)*	1.71 (1.27)	3.79 (1.68)*
Random Effect (Model 4)	Level 1	2.51 (0.68)***	2.19 (0.59)***	1.93 (0.51)***
	Level 2	2.44 (1.28)*	1.42 (0.84)^{t}	3.60 (1.55)*
Test of significance				
Reference model		195.025	188.614	182.274
Δ-2logV		181.254	171.192	177.605
χ^2 (df)		13.771 (2)***	17.422 (2)***	4.669 (2)^{t}
Effect size |d|		0.73	0.76	0.57

Notes. Model 4: Treatment Condition = 0 and Control Condition = 1. Sex: male=1

^{t} < .10, *p < .05, **p < .01, ***p < .001

Table 3

Results of the Multilevel Models for Standardized percentage of MVPA

		Unconditional model - three levels	Unconditional model- two levels	Conditional model-two levels
		b (SE)	*b (SE)*	*b (SE)*
Fixed Effect	Intercept	0.77 (2.81)	0.78 (2.81)	6.24 (13.20)
	Sex			7.04 (1.47)***
	Age			-1.46 (1.34)
	Condition			6.29 (6.18)
	Time			5.42 (1.38)***
	Time x Condition			-8.70 (2.01)***
Random Effect	Level 1	339.00 (24.57)***	349.36 (19.85)***	325.25 (18.49)***
	Level 2	10.68 (16.48)		
	Level 3	108.94 (43.79)*	109.20 (43.75)*	120.49 (48.90)*
Test of significance				
Reference model				5540.409
Δ-2logV				5516.916
χ^2 (df)				23.493 (2)***
Effect size |d|				1.33

Notes. Model 4: Treatment Condition = 0 and Control Condition = 1. Sex: male=1

****p* < .001

Supplementary material

Components of the Professional Development

The teachers in TC were invited to participate in a four 3-hours TPD program offered by a researcher specialist in PE teaching and sport psychology. The three major aims of the TPD program were to: (1) explain to the teachers that some motivating styles can be detrimental (i.e., undermine students' psychological needs) and others can be beneficial (i.e., support students' psychological needs) to students' motivation, (2) help teachers to plan their lessons by answering the question: "how can I use motivating styles that will help my students improve their effort in PE?", and (3) help teachers become more independent in the design of their PE lessons. Each of the four TPD sessions was divided in 3 parts. Within these 3 parts we took into account the six design features of effective intervention programs (Su & Reeve, 2011).

More specifically, in every TPD session, the first part was dedicated to present basic tenets of SDT, including the six dimensions of motivating style and their consequences on students' effort in PE. Further, empirical and theoretical evidence was presented to support the argument that students' apply more effort in PE when teachers support their basic psychological needs. Video footages of PE teaching were used to illustrate each component of the teachers' motivating style. We used these footages, because Tessier et al. (2010) showed that teachers implemented more need-supportive strategies after they observed their own motivating style from video footages of their own lessons. Starting from the second session in December, the participants analyzed their own video footage. This analysis required the use of the SDT framework to identify dimensions of need-supportive motivating style that the teachers had mastered and dimensions they needed to further work on.

The second part of the four sessions consisted of implementing, in the targeted sport, the strategies proposed during the first part of the session. Choosing from a learning situation (e.g., pairs building an itinerary in orienteering, portraying the same message with different part of the body in dance) covered in the notebook, the teachers gathered in pairs or triads to plan a part of their future PE lesson (e.g., warm-up). They had to discuss why they chose a particular learning situation, what their objective was, how they would administer their instructions (e.g., to the entire class *versus* small groups of students), what kind of feedback and when they would deliver it, what modifications they could add to make the situation easier or more difficult, and when to use this situation – as warm-up, as a learning task, or at the end of the lesson. Each group chose a different learning situation. Teachers were assisted in this part of the TPD session by the workshop facilitator.

In the last part of each session this plan was put into practice (i.e., teachers enacted the situation with the other teachers playing the role of the student). Each role playing lasted about 15 minutes. Following this, the participants were invited to reflect collectively on the situation drawing from the SDT framework. They were asked if the situation nurtured their students' psychological needs (e.g., did the teacher provide choices, encourage their students and give feedback, was his/she warm or show his/her concern?), and how could they improve their planning or motivating style.

As the teachers became more theoretically knowledgeable over time, the TPD sessions allocated gradually less time on the SDT principles and became more focused on the practical parts (didactic and pedagogic preparation and role-playing). During all parts of each session teachers were invited to express their beliefs, concerns or doubts about applying some of the teaching strategies. Answers to their concerns were discussed collectively. To ensure that the researcher delivered the content of the training as intended, the TPD sessions were offered using power-point slides on which the key ideas for each slide were available to the presenter

in a note format. In addition, TPD sessions were rehearsed before being presented to the teachers to ensure that sufficient time was dedicated to clearly deliver the content of each part of the training.

Based on Su and Reeve's (2011) recommendations, in addition to the TPD sessions, teachers in the TC benefited from onsite follow-up individualized guidance, which was offered by the researcher when the teachers were visited for data collection purposes. This follow-up consisted of a 10-to-15 minute session aimed at helping teachers to analyze their actual behavior using video footage, and to find alternatives that better satisfy their students' psychological needs. When teachers did not manage to identify alternatives by themselves, the researcher offered some suggestions. In some cases, when a teacher thought that a need-supportive motivating style was not the best option (e.g., when reminding students about safety rules or restoring discipline in the classroom), the researcher did not try to encourage the use of a new strategy. Teachers in CC were also visited at equivalent time points for data collection purposes, but they were not offered any follow-up sessions.

Coders' training

Training consisted of PowerPoint presentations (e.g., presentation of the SDT tenets and the observation tool), small group seminars (e.g., review of quiz and clarification of the motivating style's dimensions), and collaborative and independent coding sessions (e.g., appropriation of the marking scheme and observation schedule) and took approximately 6 hr to complete. To finish their coder training, coders were asked to rate two pilot videos using the observational instrument. These ratings were compared with a "gold standard" rating from the lead researcher. If the weighted kappa of Cohen coefficient surpassed an acceptable level of reliability (i.e., $K > 0.70$) coders began rating the main trial footage.

Coding protocol

Teachers in both conditions were videotaped using a digital camcorder, and their communications with students were recorded using a small microphone fixed on the collar of their sweatshirt. Each video of a PE lesson was divided into four equal quarters to allow for a sufficient time period on which to base the resulting potency rating. The duration of a quarter ranged from 10 minutes to 16 minutes. The coders then added the scores of the four equal quarters in order to obtain a score out of 12 for each of the six dimensions of the motivating style. We calculated the average of the scores provided by the two coders when the difference between their scores was 3 or below (out of a maximum score of 12). When the difference in the scores was above these thresholds, a third coder was invited to code the video footages. Then, coding discrepancies were discussed among all three coders until a consensus was reached.

Steps to perform multilevel analysis for linear and quadratic models.

For each dependent variable the following steps were followed.

Step1: unconditional model. In this model, no predictor was included. It served as a baseline model to examine individual variation in the outcome variable, collapsed across time.

Step 2: unconditional linear growth curve model. The variable 'time' was included as a fixed parameter/predictor. This variable represented the linear change in the group coded as 0 (i.e. the treatment condition) over time. This model examined individual changes over time in the outcome variable (i.e., how each person's rate of change deviates from the true rate of change of the population)

Step 3: in this model we tested for group differences in linear change over time. Thus the experimental condition (a dummy variable) and the interaction term of time × condition were added as predictors (see model 3). In these analyses we controlled for the effects of teachers' gender and their teaching experience.

Step 4: the variable time² and the interaction time² × condition (model 4) were added as predictors. These predictors tested for group differences in curvilinear change over time

At step 2 and onwards, the -2 the log likelihood was used to compare the model tested at that step and the immediately previous model. The more the -2 log likelihood value decreased, the better the model fit.

Table 1

Descriptive Statistics and Weighted Kappa for Motivating Styles and PA Levels

	Variables	Time 1			Time 2			Time 3			Time 4		
		M	*SD*	K[a]	*M*	*SD*	K[a]	*M*	*SD*	K[a]	*M*	*SD*	K[a]
Control condition	1. Autonomy Support	3.79	2.86	0.96	3.83	2.02	0.91	5.00	2.65	0.94	4.13	3.79	0.95
	2. Control	3.50	3.16	0.82	6.08	3.20	0.90	3.00	1.00	0.89	3.63	3.40	0.95
	3. Structure	5.43	1.62	0.87	7.25	1.41	0.85	6.17	2.25	0.95	5.25	2.22	1.00
	4. Chaos	2.36	3.42	1.00	1.33	1.21	0.96	4.50	4.27	0.93	4.25	3.86	0.95
	5. Involvement	6.14	2.87	0.77	4.92	2.40	1.00	6.00	5.57	0.95	6.00	1.35	0.93
	6. Hostility	3.50	3.30	0.84	6.25	2.56	0.91	2.33	2.30	0.88	2.25	2.63	1.00
	7. Percentage of MVPA	65	20		68	21		43	22		-	-	
	8. Centered percentage of MVPA	0.48	20		-0.46	21		-5.93	22		-	-	
Treatment condition	1. Autonomy Support	4.20	2.39		6.43	2.99		8.42	1.43		4.50	1.32	
	2. Control	2.80	2.49		2.86	2.81		3.25	2.36		2.80	1.10	
	3. Structure	5.10	2.30		8.21	2.98		9.50	2.35		6.80	1.64	
	4. Chaos	1.80	1.64		1.79	2.71		1.33	1.51		2.80	2.77	
	5. Involvement	4.80	1.64		6.43	3.31		6.75	2.27		6.80	1.92	
	6. Hostility	2.20	1.64		1.71	1.2		2.25	2.51		2.60	2.30	
	9. Percentage of MVPA	64	18		69	25		53	18		-	-	
	10. Centered percentage of MVPA	-0.66	18		0.53	25		3.85	18		-	-	

Notes. MVPA= moderate to vigorous PA

[a] Weighted kappa was analyzed based on data from the whole sample

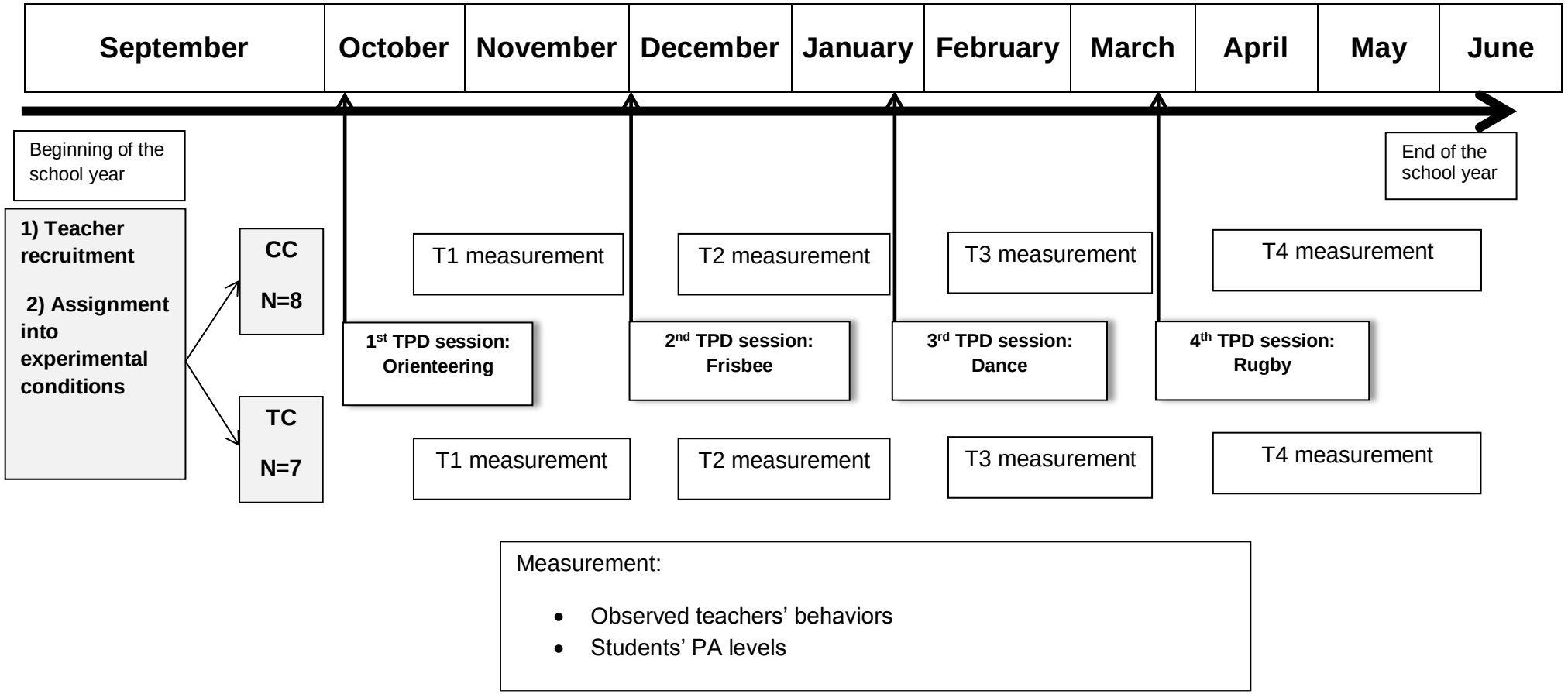

Figure 1. Procedural timeline for the need-supportive teacher training intervention and data collection

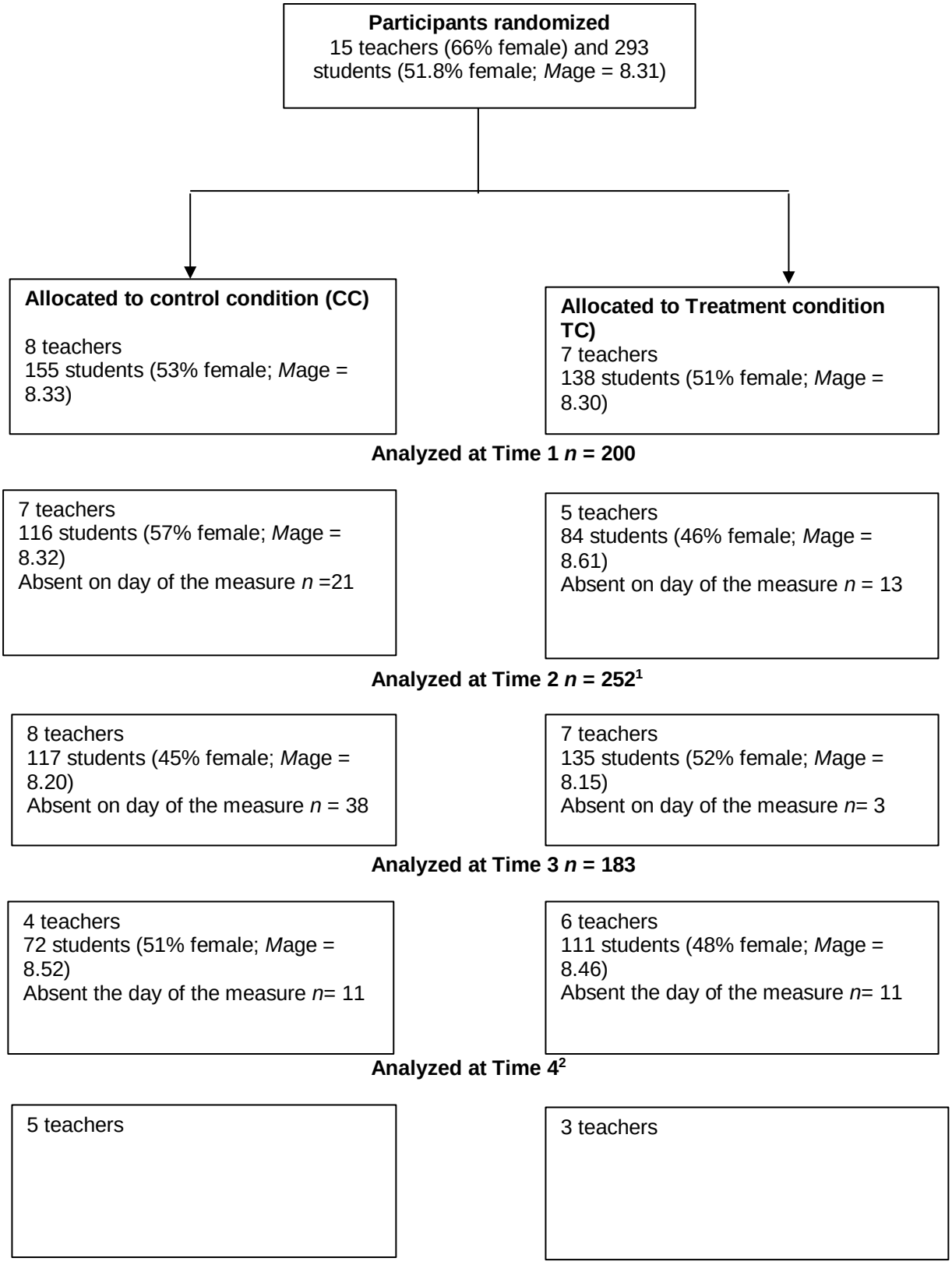
Participants randomized
15 teachers (66% female) and 293 students (51.8% female; *M*age = 8.31)
Allocated to control condition (CC)
8 teachers
155 students (53% female; *M*age = 8.33)
Allocated to Treatment condition TC)
7 teachers
138 students (51% female; *M*age = 8.30)
Analyzed at Time 1 *n* = 200
7 teachers
116 students (57% female; *M*age = 8.32)
Absent on day of the measure *n* =21
5 teachers
84 students (46% female; *M*age = 8.61)
Absent on day of the measure *n* = 13
Analyzed at Time 2 *n* = 252[1]
8 teachers
117 students (45% female; *M*age = 8.20)
Absent on day of the measure *n* = 38
7 teachers
135 students (52% female; *M*age = 8.15)
Absent on day of the measure *n*= 3
Analyzed at Time 3 *n* = 183
4 teachers
72 students (51% female; *M*age = 8.52)
Absent the day of the measure *n*= 11
6 teachers
111 students (48% female; *M*age = 8.46)
Absent the day of the measure *n*= 11
Analyzed at Time 4[2]
5 teachers
3 teachers

Figure 2. Flow diagram

[1] The sample was larger at time 2 than at time 1 is because some lessons were cancelled at time 1 due to weather conditions and teacher ill health; it was not possible to reschedule our visits later in the first period

[2]There are no student data at this wave because most teachers did not consent to their students wearing an accelerometer (see text for an explanation)

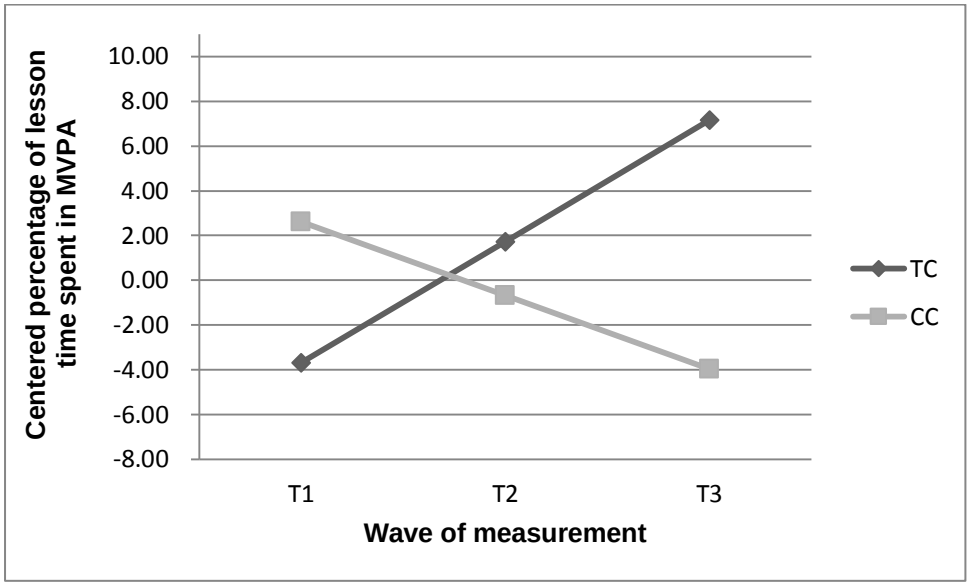

Figure 3. Interaction time × condition for centered percentage of MVPA. TC= treatment condition; CC= control condition; MVPA= moderate to vigorous physical activity

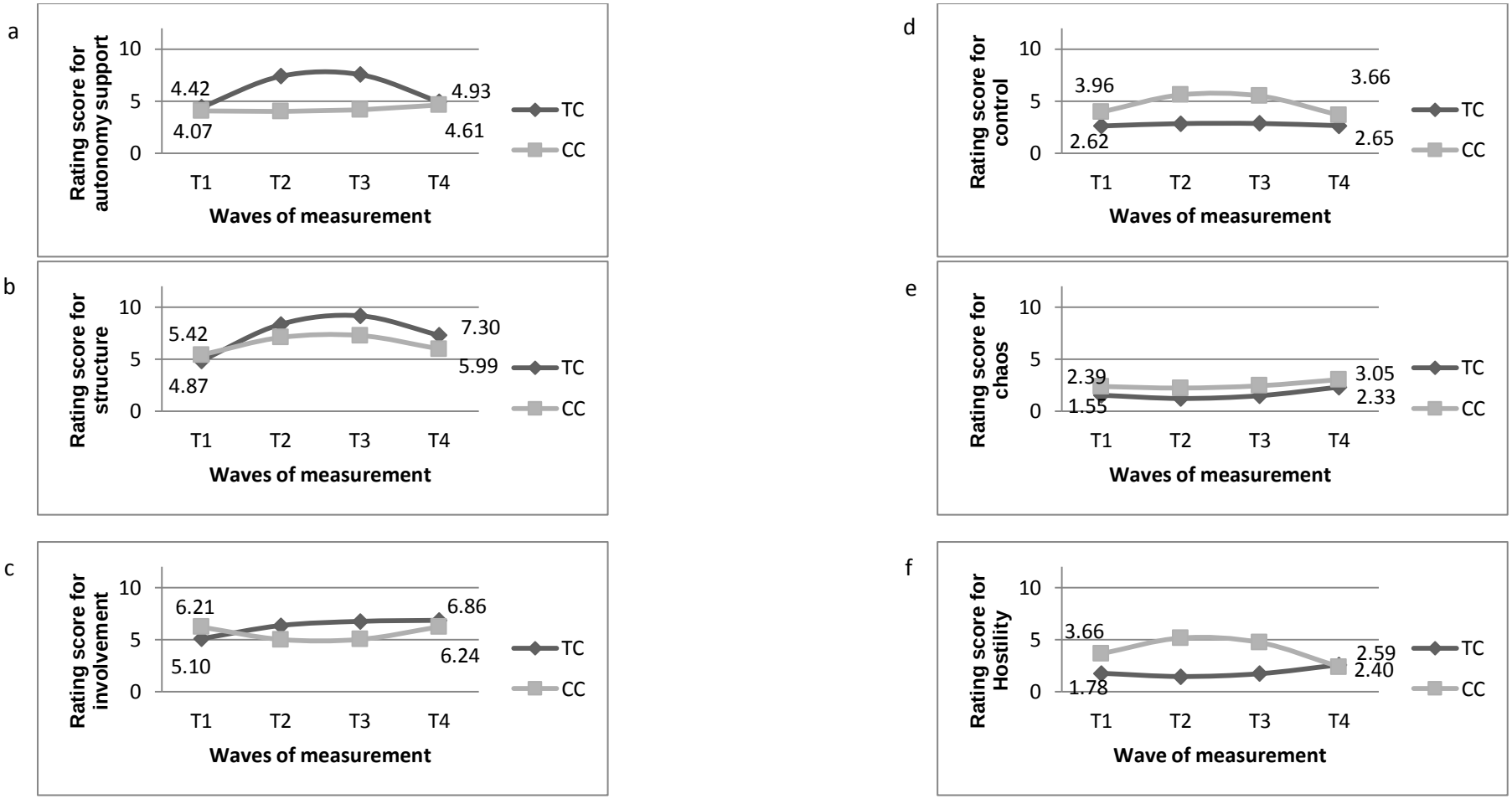

Figure 4. Interaction time² × condition for autonomy support, structure, involvement, control, chaos, and hostility. TC= treatment condition; CC= control condition

Section/Topic	Item No	Standard Checklist item	Extension for cluster designs	Page No *
Title and abstract				
	1a	Identification as a randomised trial in the title	Identification as a cluster randomised trial in the title	p.1
	1b	Structured summary of trial design, methods, results, and conclusions (for specific guidance see CONSORT for abstracts)	See table 2	p.2
Introduction				
Background and objectives	2a	Scientific background and explanation of rationale	Rationale for using a cluster design	Introduction pp. 3-8 and Method p.11
	2b	Specific objectives or hypotheses	Whether objectives pertain to the the cluster level, the individual participant level or both	Introduction pp. 8-10
Methods				
Trial design	3a	Description of trial design (such as parallel, factorial) including allocation ratio	Definition of cluster and description of how the design features apply to the clusters	Method p.11-12
	3b	Important changes to methods after trial commencement (such as eligibility criteria), with reasons		NA
Participants	4a	Eligibility criteria for participants	Eligibility criteria for clusters	Method p. 10-11
	4b	Settings and locations where the data were collected		Method p. 10
Interventions	5	The interventions for each group with sufficient details to allow replication, including how and when they were actually administered	Whether interventions pertain to the cluster level, the individual participant level or both	Method pp. 11-13 and supplementary material
Outcomes	6a	Completely defined pre-specified primary and secondary outcome measures, including how and when they were assessed	Whether outcome measures pertain to the cluster level, the individual participant level or both	Measures pp.13-15
	6b	Any changes to trial outcomes after the trial		Results p.17

		commenced, with reasons		
Sample size	7a	How sample size was determined	Method of calculation, number of clusters(s) (and whether equal or unequal cluster sizes are assumed), cluster size, a coefficient of intracluster correlation (ICC or *k*), and an indication of its uncertainty	Method p.10-11
	7b	When applicable, explanation of any interim analyses and stopping guidelines		NA
Randomisation:				
Sequence generation	8a	Method used to generate the random allocation sequence		Method p.11
	8b	Type of randomisation; details of any restriction (such as blocking and block size)	Details of stratification or matching if used	Method p.11
Allocation concealment mechanism	9	Mechanism used to implement the random allocation sequence (such as sequentially numbered containers), describing any steps taken to conceal the sequence until interventions were assigned	Specification that allocation was based on clusters rather than individuals and whether allocation concealment (if any) was at the cluster level, the individual participant level or both	Method p.11
Implementation	10	Who generated the random allocation sequence, who enrolled participants, and who assigned participants to interventions	Replace by 10a, 10b and 10c	
	10a		Who generated the random allocation sequence, who enrolled clusters, and who assigned clusters to interventions	Method p.12
	10b		Mechanism by which individual participants were included in clusters for the purposes of the trial (such as complete enumeration, random sampling)	Method p.11

	10c		From whom consent was sought (representatives of the cluster, or individual cluster members, or both), and whether consent was sought before or after randomisation	Method p.11
Blinding	11a	If done, who was blinded after assignment to interventions (for example, participants, care providers, those assessing outcomes) and how		Measures p.13
	11b	If relevant, description of the similarity of interventions		Measures p.13-14
Statistical methods	12a	Statistical methods used to compare groups for primary and secondary outcomes	How clustering was taken into account	Data analysis p.15-16
	12b	Methods for additional analyses, such as subgroup analyses and adjusted analyses		NA
Results				
Participant flow (a diagram is strongly recommended)	13a	For each group, the numbers of participants who were randomly assigned, received intended treatment, and were analysed for the primary outcome	For each group, the numbers of clusters that were randomly assigned, received intended treatment, and were analysed for the primary outcome	Method p.11 and Supplementary material: flow diagram
	13b	For each group, losses and exclusions after randomisation, together with reasons	For each group, losses and exclusions for both clusters and individual cluster members	Results p.17
Recruitment	14a	Dates defining the periods of recruitment and follow-up		Supplementary material: procedural timeline
	14b	Why the trial ended or was stopped		NA
Baseline data	15	A table showing baseline demographic and clinical	Baseline characteristics for the individual and cluster levels as applicable for each group	NA

		characteristics for each group		
Numbers analysed	16	For each group, number of participants (denominator) included in each analysis and whether the analysis was by original assigned groups	For each group, number of clusters included in each analysis	Supplementary material: flow diagram
Outcomes and estimation	17a	For each primary and secondary outcome, results for each group, and the estimated effect size and its precision (such as 95% confidence interval)	Results at the individual or cluster level as applicable and a coefficient of intracluster correlation (ICC or k) for each primary outcome	Results pp.17-21
	17b	For binary outcomes, presentation of both absolute and relative effect sizes is recommended		NA
Ancillary analyses	18	Results of any other analyses performed, including subgroup analyses and adjusted analyses, distinguishing pre-specified from exploratory		NA
Harms	19	All important harms or unintended effects in each group (for specific guidance see CONSORT for harms)		NA
Discussion				
Limitations	20	Trial limitations, addressing sources of potential bias, imprecision, and, if relevant, multiplicity of analyses		Discussion p.25-27
Generalisability	21	Generalisability (external validity, applicability) of the trial findings	Generalisability to clusters and/or individual participants (as relevant)	Discussion pp.21-27
Interpretation	22	Interpretation consistent with results, balancing benefits and harms, and considering other relevant evidence		Discussion pp.20-27
Other information				

Registration	23	Registration number and name of trial registry		NA
Protocol	24	Where the full trial protocol can be accessed, if available		Supplementary material
Funding	25	Sources of funding and other support (such as supply of drugs), role of funders		NA

[1] In the French education system, elementary school teachers have to teach all subjects, including PE. However, within their pre-service program, among the 400 hours for courses, only 40 hours are allocated to teaching PE. Thus, French elementary school teachers are not specialist in PE teaching. In contrast, secondary PE teachers attend programs focused on the pedagogical knowledge needed to teach the French PE curriculum, hence they are PE specialists.