



Performance grading and motivational functioning and fear in physical education: A self-determination theory perspective[☆]



Christa Krijgsman^{a,b,*}, Maarten Vansteenkiste^c, Jan van Tartwijk^a, Jolien Maes^b, Lars Borghouts^d, Greet Cardon^b, Tim Mainhard^a, Leen Haerens^b

^a Department of Education, Utrecht University, The Netherlands

^b Department of Movement and Sport Sciences, Ghent University, Belgium

^c Department of Developmental, Personality and Social Psychology, Ghent University, Belgium

^d School of Sport Studies, Fontys University of Applied Sciences, Eindhoven, The Netherlands

ARTICLE INFO

Article history:

Received 25 February 2016

Received in revised form 17 December 2016

Accepted 27 March 2017

Keywords:

Need satisfaction

Need frustration

Motivation

Assessment

Fear

ABSTRACT

Grounded in self-determination theory, the present study examines the explanatory role of students' perceived need satisfaction and need frustration in the relationship between performance grading (versus non-grading) and students' motivation and fear in a real-life educational physical education setting. Grading consisted of teacher judgments of students' performances through observations, based on pre-defined assessment criteria. Thirty-one classes with 409 students ($M_{age} = 14.7$) from twenty-seven Flemish (Belgian) secondary schools completed questionnaires measuring students' perceived motivation, fear and psychological need satisfaction and frustration, after two lessons: one with and one without performance grading. After lessons including performance grading, students reported less intrinsic motivation and identified regulation, and more external regulation, amotivation and fear. As expected, less need satisfaction accounted for (i.e., mediated) the relationship between performance grading and self-determined motivational outcomes. Need frustration explained the relationship between performance grading and intrinsic motivation, as well as less self-determined motivational outcomes. Theoretical and practical implications are discussed.

© 2017 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

Using grades to assess students' performance is an integral part of educational systems around the globe (Ames, 1992; Lingard, 2010; Strain, 2009). The motivational impact of grading is likely to depend on its functional significance (Vansteenkiste, Ryan, & Deci, 2008). When students predominantly perceive a grading event as a judgment of their performance, rather than as a way of receiving information about their learning process, this may come at a motivational price (Ames, 1992; Amrein & Berliner, 2002; Ryan & Brown, 2005). Students' focus on performing well to obtain good grades may then undermine their interest and 'love of learning' (Butler, 1987; Butler & Nisan, 1986; Jones, 2007; Pulfrey, Darnon, & Butera, 2013). Moreover, students may start to avoid looking bad in front of their teachers or peers, which

may lead to fear of failure and feelings of incompetence when grades are inferior (Elliot & McGregor, 1999; McDonald, 2001; Ryan & Weinstein, 2009). Using a within-person design, the present research investigated whether students' motivational functioning, fear and need-based experiences varied as a function of whether they were graded or not during their real-life physical education (PE) classes (i.e., ecologically valid setting). Moreover, extending past work, we addressed the processes (i.e., need-based experiences) underlying the hypothesised motivational and fear differences between a grading and non-grading class. Because the functional significance of the grading was primarily evaluative and judgmental of student's performance, we refer to this type of grading as 'performance grading'.

1.1. Grading in physical education

As in many other countries, in Flanders (Belgium), PE students are regularly assessed throughout the school year. Functions of assessment in PE (as in academic courses) can be positioned on a continuum from 'performance-based assessment' (i.e., quality judgment of students' performance) to 'informational assessment' (i.e., specifying learning progress and constructing the way forward; López-Pastor, Kirk, Lorente-Catalán, MacPhail, & Macdonald, 2012; Tunstall & Gipps,

[☆] This research was supported by the Netherlands Organisation for Scientific Research (Grant 023.004.015).

* Corresponding author: Heidelberglaan 1, 3508 TC Utrecht, The Netherlands.

E-mail addresses: c.a.krijgsman@uu.nl, Christa.Krijgsman@UGent.be (C. Krijgsman), Maarten.Vansteenkiste@UGent.be (M. Vansteenkiste), j.vantartwijk@uu.nl (J. van Tartwijk), Jolien.Maes@UGent.be (J. Maes), l.borghouts@fontys.nl (L. Borghouts), Greet.Cardon@UGent.be (G. Cardon), m.t.mainhard@uu.nl (T. Mainhard), Leen.Haerens@UGent.be (L. Haerens).

1996). In Flanders (Belgium), PE students are often exposed to a performance-based assessment system. Students' performance is commonly rated with the grades 1 to 10. The grades '1' to '4' designate an insufficient performance, the grades '5' to '7' describe a sufficient performance, and the grades '8' to '10' describe good to excellent performances (i.e., a 'multiple grades system'; Barenberg & Dutke, 2013, p.122).

While awarding performance-based grades in PE, teachers typically use criterion referenced grading (i.e., how well do students perform relative to criteria; Pulfrey, Buchs, & Butera, 2011; Redelius & Hay, 2012) and norm referenced grading (how well students perform relative to others; Chan, Hay, & Tinning, 2011; Elliot & Moller, 2003; Johnson, Prusak, & Pennington, 2011). Frequently used methods are teacher judgments based on observations with (Borghouts, Slingerland, & Haerens, 2016; Svennberg, Meckbach, & Redelius, 2014) or without (Annerstedt & Larsson, 2010; Hay & Macdonald, 2008; Svennberg et al., 2014) explicitly communicating criteria. Irrespective of the type of grading that students are submitted to, or which combination of grading systems the teacher employs, assessing performance through the use of a multiple grades system conveys information, which allows (and in fact mostly triggers) students to compare their performance with other students. Moreover, students in Flanders (Belgium) receive a report card at the end of each semester, which contains the average grades for PE along with other subjects (European Commission/EACEA/Eurydice, 2013). This report card again allows students to directly compare performances. It is therefore argued that performance-based grades stimulate normative and social comparison (Ames, 1992; Elliot & Moller, 2003). Such social comparison (Ames, 1992) might be further fostered by the 'visibility' of performance during PE lessons (Annerstedt & Larsson, 2010; Johnson et al., 2011; Redelius & Hay, 2012), and may come with a motivational cost.

1.2. Self-determination theory and performance grading

1.2.1. Motivational differences

According to SDT, depending on whether the performance grading is perceived to be more evaluative and judgmental or informational and helpful, different types of motivation are likely to be engendered. A refined taxonomy of motives is discerned within SDT, with some of them being more autonomous and others more controlled in nature (Deci & Ryan, 2000; Vansteenkiste, Lens, & Deci, 2006). Students are said to display autonomous regulation during a PE class when they find their class to be enjoyable and interesting (i.e., intrinsic motivation) or value its personal benefits (i.e., identified regulation). In contrast, students are controlled motivated when they put effort in their PE class to please their teacher, to obtain good grades, or to avoid criticism (i.e., external regulation). Interestingly, students may not only be externally pressured, but could also pressure themselves to do well (i.e., introjected regulation), for instance by buttressing their activity engagement with feelings of guilt and contingent self-worth. While students are – quantitatively speaking – motivated when they display either autonomous or controlled motivation, amotivation within SDT reflects a lack of motivation. Specifically, amotivated students typically invest a minimum amount of effort in PE classes because they experience incapability to perform activities, or because they do not experience a personal value (Deci & Ryan, 2000).

Dozens of previous studies have indicated that autonomous motivation, relative to controlled motivation and amotivation, relates to a host of desirable outcomes (see Ntoumanis & Standage, 2009 for an overview). To illustrate, autonomous motivation is predictive of students' observed engagement (Aelterman et al., 2012) and rated performance (Vansteenkiste, Simons, Lens, Sheldon, & Deci, 2004), whereas controlled motivation and amotivation relate to undesirable outcomes, including boredom (Ntoumanis, 2001), low engagement (Aelterman et al., 2012), and fear of exams and test situations (Schaffner & Schiefele, 2007).

Further, a number of studies have indicated that these different types of motivation get differentially activated under grading versus non-grading circumstances. For instance, experimental research showed that grading, particularly when students experience it as a judgment of their performance, results in lower levels of intrinsic motivation (Butler, 1987; Butler & Nisan, 1986; Grolnick & Ryan, 1987; Johnson et al., 2011; Pulfrey et al., 2011) and identified regulation (Johnson et al., 2011; Pulfrey et al., 2011). Furthermore, two studies found external regulation (Grolnick & Ryan, 1987; Johnson et al., 2011) and amotivation (Johnson et al., 2011) to increase in situations where performance-based grading takes place. Yet, to the best of our knowledge, no previous study specifically examined the relationship between performance grading and introjected regulation. Although it seems rather self-evident that students are more externally regulated during an evaluative grading class, the question remains whether they equally apply such pressure to their own functioning. Presumably, because performance grading 'awakens' students' ego, they may display more introjected regulation as well.

1.2.2. Explanatory processes: need-based experiences

While the motivational correlates of performance grading are fairly well documented in the literature, less is known about the processes underlying these effects (but see Pulfrey et al., 2013). To predict the motivational impact of performance grading, from a SDT-account, the critical question is whether the grading impacts on individuals' psychological need-based experiences. Three psychological needs have been discerned, that is, the need for autonomy, relatedness, and competence (Deci & Ryan, 2000). Specifically, need satisfaction refers to students' experience of volition and self-endorsement (i.e., need for autonomy), their feeling of connection and mutual care (i.e., need for relatedness) and their experience of effectiveness (i.e., need for competence). Dozens of studies have indicated that the satisfaction of these needs contributes to individuals' autonomous motivation, and their engagement and growth in the classroom (Niemiec & Ryan, 2009).

While the satisfaction of these needs has received considerable attention, it is only more recently that the notion of need frustration, which may particularly be useful in the context of grading, has been researched more intensively (Bartholomew, Ntoumanis, Ryan, Bosch, & Thøgersen-Ntoumani, 2011a; Bartholomew, Ntoumanis, Ryan, & Thøgersen-Ntoumani, 2011b; Haerens, Aelterman, Vansteenkiste, Soenens, & van Petegem, 2015). Need frustration deserves attention by its own right because – theoretically speaking – the absence of need satisfaction does not necessarily denote the presence of need frustration (Vansteenkiste & Ryan, 2013). Indeed, for need frustration to occur, a more active thwarting of individuals' needs is required. Specifically, need frustration refers to feelings of pressure and internal conflict (i.e., autonomy frustration), rejection and disrespect (i.e., relatedness frustration), or feelings of failure and inadequacy (i.e., competence frustration). The distinction between need satisfaction and frustration is critical as unfulfilled needs (i.e., low need satisfaction) may not relate as robustly to malfunctioning as frustrated needs may. A metaphor (Vansteenkiste & Ryan, 2013, p.265) may help to account for this assumption: 'If plants do not get sunshine and water (i.e., resulting in low need satisfaction), they will fail to grow and will die over time; yet, if salted water is thrown on plants (i.e., eliciting need frustration), they will wither more quickly.' Thus, whereas low need satisfaction is likely to yield motivational costs over time, high need frustration will accelerate negative motivational processes. Congruent with this assumption, past research has found need satisfaction to be predictive of autonomous motivation (Haerens et al., 2015), engagement (Jang, Kim, & Reeve, 2016) and well-being (Bartholomew et al., 2011a), while need frustration relates to controlled motivation and amotivation (Haerens et al., 2015), disengagement (Jang et al., 2016) and ill-being (Bartholomew et al., 2011a). Such findings have been documented using cross-sectional, longitudinal and diary designs (van der Kaap-Deeder et al., 2016).

The question whether performance grading relates to individuals' need-based experiences has received little attention (but see Pulfrey et al., 2013). It is possible that when exposed to grading, especially if the grading is evaluative and judgmental, students might not only experience a lack of choice or freedom (i.e., low autonomy satisfaction), they may also feel pressured to perform well (i.e., high autonomy frustration). Likewise, students might not only experience a sense of disconnection to others (i.e., low relatedness satisfaction), they might also feel rejected by others when anticipating (reactions to) a lower grade (i.e., high relatedness frustration). In a similar vein, students might not only think they will not be able to reach the criteria (i.e., low competence satisfaction), in some situations they might even feel like a failure (i.e., high competence frustration), particularly if they receive bad grades despite their efforts. Consistent with these prior assumptions, it was indicated that autonomy satisfaction and competence satisfaction accounted for the link between performance grading and task interest (i.e., intrinsic motivation; Pulfrey et al., 2013). Yet, whether and how experienced need satisfaction and frustration vary as a function of performance grading and whether these need-based experiences can account for the hypothesised link between performance grading and the broad spectrum of students' motivational functioning and fear has not received any attention so far.

1.3. The present study

Grounded in self-determination theory, the present study, conducted in an ecological valid setting (i.e., during authentic lessons), addressed the following research questions: (1) Do students' display different motivational functioning, fear and need-based experiences in a PE lesson in which performance grading is applied compared with a lesson in which no performance grading is applied and (2) can differences in motivational functioning and fear be accounted for by differences in experienced need satisfaction and frustration across both lessons? While performance grading in PE might be considered low-stake, we posit nevertheless that participating in grading activities in PE might be associated with more fear and a different pattern of motivational functioning and need-based experiences. We formulated the following two hypotheses.

First, based on previous research (Grolnick & Ryan, 1987; Johnson et al., 2011; Putwain & Best, 2011) we hypothesised that students would report lower levels of intrinsic motivation, identified regulation and need satisfaction, and higher levels of introjected regulation, external regulation, amotivation, fear and need frustration, when being exposed to a performance-based grading class versus a non-grading class (see also Butler & Nisan, 1986; Pulfrey et al., 2011).

Second, we investigated the explanatory (i.e., mediating) role of students' experiences of need satisfaction and need frustration in the relationship between performance grading (versus non-grading) and the hypothesised differences in motivation and fear. Given that this question has not received any prior attention (but see Pulfrey et al., 2013), we were open for the possibility that performance grading may come with low need satisfaction or a combination of low need satisfaction and high need frustration. To illustrate, performance grading may reduce feelings of choice or freedom (i.e., low autonomy satisfaction), and may simultaneously increase students' pressure and stress to perform well (i.e., autonomy frustration). If differences in need frustration would surface, they may help to account for why grading versus non-grading relates to students' higher levels of introjected regulation, external regulation, amotivation and fear.

2. Method

2.1. Participants

A convenience sample of thirty-nine PE teachers (24 males; 61.5%) and 724 students (399 boys; 55.1%, $M_{age} = 14.7 \pm 0.94$) from 39 schools

in Flanders (Belgium) participated in the study. Of all 724 participating students, 315 students did not have complete measures, and therefore these students were excluded from the analyses, resulting in a final sample of twenty-seven schools with thirty-one PE teachers (21 males; 67.7%) and 409 students (response rate = 57%, 222 boys; 54.3%, $M_{age} = 14.7 \pm 1.00$). In Flanders (Belgium), the gender formation of PE classes (i.e., mixed gender grouped or single gender grouped) depends on the districts in which schools are located. In the present sample, of all 31 classes, 14 classes (45.2%) were mixed gender grouped and 17 classes (54.8%) were single gender grouped (11 male classes; 64.7%). Class sizes ranged from eight to thirty students per class ($M = 17 \pm 5.18$). All students attended secondary education: 46.2% of the students attended academic education, 31.1% technical education and 22.7% vocational education.

2.1.1. Ethical considerations

All participating teachers and their principals gave informed consent to their participation in the current study. With the exception of eleven parents, all parents gave informed consent for their child's participation. All participants were assured that responses were treated confidentially. The Ethical Committee of Ghent University approved the study protocol.

2.2. Procedure

For the purposes of the present study, teachers were asked to give their lessons as planned. In Flanders (Belgium), PE is a compulsory subject in secondary schools for at least two 50-min lessons each week. These two 50-min lessons are sometimes combined into one single 100-min lesson. The research leader plus a team of research assistants collected the data. Students filled out a set of questionnaires during the last 15 min of two lessons out of a sequence of lessons on one specific topic (e.g., four basketball lessons). The first measurement took place at the end of the first or second lesson of the series of lessons: a lesson in which students were not graded. The second measurement took place in the final lesson of the series of lessons: a lesson in which students received a performance grade. Students were aware of the fact that they were graded during this specific lesson. The time frame between both measurements was in most classes one to three weeks. No manipulations were made to the normal procedure in the lessons, with the exception of filling out the questionnaires at the end of both lessons.

To understand how students in the present sample were assessed, data was collected with two different types of measurements. First, teachers were questioned about their grading practices by means of open questions. In these questionnaires, teachers indicated that it was usual to grade their students on a specific lesson topic in a relatively short period of time. For most teachers in this sample it was common practice to teach about three to four consecutive lessons on one subject (e.g., four lessons of basketball) with grading taking place in the last lesson. In the lessons in which grading took place, teachers graded students' motor skills on the same specific subject (e.g., grading the lay-up as a basketball technique in the final lesson of four lessons in which students had practiced the lay-up) to obtain a qualification of students' performance. Second, 30 teachers ($n = 1$ missing) were filmed in the lessons in which grading took place. Observations of these lessons provided a good indication of the actual grading practices. It was clear from the videos that all grading lessons had the purpose of qualifying students' performance at the end of a learning process. Except for one teacher (for whom we could not verify from video or questionnaire whether students' performance was qualified by means of a grade), all teachers assessed students by means of a grade. Video observations indicated that, while awarding performance-based grades in PE, the majority of the teachers in our sample informed the students about assessment-criteria. These criteria were designed and used to measure product performance (i.e., purely measuring students' performance at

the end of the learning process). After communicating these criteria, with the exception of one teacher, all teachers awarded performance grades based on their own observations and judgments (i.e., one teacher used peer assessment). Almost half of the students received their grade in the grading lesson itself. Other students had to retrieve the grade at a later moment from a digital system. Independent of the method used for grading, largely all students worked in small groups while being graded and hardly any assessment tools were used, such as videos or photos for observation.

2.3. Measures

2.3.1. Motivational regulations

Insights into students' motivational regulations towards the last PE class were obtained by use of the Behavioural Regulations in Physical Education Questionnaire (BRPEQ; Aelterman et al., 2012) in a similar way as it was done in previous research (Aelterman et al., 2012; Haerens et al., 2015). Table 1 reports on the typical items, reliability and number of items per scale and per measurement occasion. Students responded to all items on a 5-point Likert scale ranging from 'not at all true for me' to 'very true for me'. Factorial validity was examined by modeling a confirmatory factor analysis (CFA) per time point, performed with Mplus (version 7.4; Muthén & Muthén, 2015). The time point 1 model fitted the data well (for recommendations see Hu & Bentler, 1999; Kline, 2011), $\chi^2(142) = 422.32$, $p < 0.001$, RMSEA = 0.07, CFI = 0.92 and SRMR = 0.06. All indicator loadings were above 0.61, $p < 0.001$. The time point 2 model fitted the data reasonably well, $\chi^2(142) = 568.63$, $p < 0.001$, RMSEA = 0.09, CFI = 0.90 and SRMR = 0.06. All indicator loadings were above 0.61, $p < 0.001$.

2.3.2. Fear

Students' level of fear was measured by means of the subscale 'fear' of the Learning And Study Strategies Inventory (LASSI; Weinstein, 1987), adapted to the context of PE. Table 1 reports on the typical items, reliability and number of items per scale, and per time point. Students responded to all items on a 5-point Likert scale ranging from 'not at all true for me' to 'very true for me'. Although the RMSEA indicated some distance between the theoretical model and the data, overall, as indicated by the CFI and SRMR, the time point 1 model fitted the data reasonably well, $\chi^2(9) = 56.40$, $p < 0.001$, RMSEA = 0.12, CFI = 0.96 and SRMR = 0.04. All indicator loadings were above 0.69, $p < 0.001$. The time point 2 model also fitted the data reasonably well, $\chi^2(9) = 56.98$, $p < 0.001$, RMSEA = 0.12, CFI = 0.97 and SRMR = 0.03. All indicator loadings were above 0.68, $p < 0.001$.

2.3.3. Need satisfaction and frustration

Students' perceived autonomy, relatedness and competence satisfaction and frustration during the last PE lesson were assessed by the Basic Psychological Need Satisfaction and Frustration Scale (BPNSFS; Chen et al., 2015). Table 1 reports on the typical items, reliability and number of items per scale and per measurement. Students responded to all items on a 5-point Likert scale ranging from 'not at all true for me' to 'very true for me'. For the purpose of the present research, small modifications were made to the original BPNSFS in order to adjust the questionnaire to the PE context. The items were modelled as indicators of six first order factors (autonomy satisfaction, autonomy frustration, relatedness satisfaction, relatedness frustration, competence satisfaction, competence frustration) that, in turn, served as indicators for two higher order factors (i.e., need satisfaction and need frustration). The time point 1 higher order model fitted the data reasonably well, $\chi^2(245) = 632.94$, $p < 0.001$, RMSEA = 0.07, CFI = 0.90, SRMR = 0.08. All indicator loadings were above 0.44, $p < 0.001$. The time point 2 higher order model also fitted the data reasonably well, $\chi^2(245) = 829.57$, $p < 0.001$, RMSEA = 0.08, CFI = 0.89, SRMR = 0.08. All indicator loadings were above 0.50, $p < 0.001$. More detailed information (i.e., all scales and subscales, factor loadings of individual items per time point) on the present study's factorial validity is presented as supplementary online data.

2.4. Plan of analysis

Given the nested structure of the data (measurements within students within classes), multilevel regression analyses were executed for all main analyses, using MLwiN version 2.30 (Rasbash, Steele, Browne, & Goldstein, 2014). When executing the main analyses, we controlled for the contextual variables gender (De Meyer et al., 2014; Johnson et al., 2011) and lesson topic (i.e., categorised as individual sports; artistic sports, and interactive sports; Aelterman et al., 2012; Guay et al., 2010) because these variables might affect students' quality of motivation, feelings of fear and need-based experiences.

Prior to the main analyses, dropout analyses, using multilevel regression analyses, were performed to examine differences between students who dropped out and those who remained in the study. Also prior to the main analyses and using multilevel regression analyses, baseline variance components models (Rasbash et al., 2014) or intercept-only models (Hox, 2010) were established for all variables in our study, with only an intercept and no explanatory variables (i.e., Model 0). As class and school level showed overlap, a three-level model (measurement, student, class) better matched the data when compared with a four-level model (measurement, student, class, school). As such, data were treated as a three-level model, in which measurements were

Table 1
Overview of the scales, number of items per scale, Cronbach's alphas per time point and example items.

Scale	N items	α		Example item
		T0	T1	
BRPEQ				<i>Using the stem</i> <i>I putted effort in the last PE class because...</i>
Intrinsic motivation	4	0.90	0.86	... I enjoyed this PE class
Identified regulation	4	0.79	0.79	... I found this PE class personally meaningful
Introjected regulation	3	0.69	0.79	... I would have felt guilty if I didn't
External regulation	4	0.78	0.90	... because I felt the pressure of others to participate in this PE class
Amotivation	4	0.80	0.87	I thought this PE class was actually a waste of time
Based upon LASSI				<i>During the last PE class...</i>
Fear	6	0.88	0.92	I thought about how bad I performed in comparison to other students
BPNSFS				<i>During the last PE class...</i>
Autonomy satisfaction	4	0.72	0.82	... I felt a sense of choice and freedom in the tasks I was participating in
Autonomy frustration	4	0.79	0.86	... I felt pressured to do certain tasks
Relatedness satisfaction	4	0.76	0.80	... I felt close and connected with other people who are important to me
Relatedness frustration	4	0.84	0.89	... I felt that people who are important to me were cold and distant towards me
Competence satisfaction	4	0.69	0.77	... I felt that I can successfully complete difficult tasks
Competence frustration	4	0.85	0.89	... I felt disappointed with many of my performances

Note. BRPEQ; Behavioural Regulations in Physical Education Questionnaire, LASSI; Learning And Study Strategies Inventory, BPNSFS; Basic Psychological Need Satisfaction and Frustration Scale.

The baseline variance components models showed a significant difference from zero in variance at class, student and measurement level (see Table 2, Model 0), for each of the motivational outcomes, fear and need-based experiences. Variance situated at the class level ranged between 6.61% for identified regulation ($\chi^2 = 4.45, df = 1, p \leq 0.05$) and 22.46% for need frustration ($\chi^2 = 10.95, df = 1, p \leq 0.001$). Variance at the student level ranged between 21.24% for external regulation ($\chi^2 = 24.13, df = 1, p \leq 0.001$) and 49.70% for identified regulation ($\chi^2 = 83.59, df = 1, p \leq 0.001$). Variance situated at the measurement level ranged between 43.69% for identified regulation ($\chi^2 = 202.13, df = 1, p \leq 0.001$) and 60.02% for external regulation ($\chi^2 = 200.56, df = 1, p \leq 0.001$).

3.2. The main analyses: motivational experiences as a function of performance grading

The first part of the main analyses was executed to answer our first research question. The predictor variable ‘grading lesson’ plus the covariates (i.e., gender and lesson theme) were added to the models examining students’ motivation, fear and need-based experiences (see Table 2, Model 1). Except for introjected regulation ($\Delta\chi^2(4) = 8.34, p = 0.08$), adding ‘grading lesson’ and the covariates to the model improved the model for all variables, as the iterated generalised least squares (IGLS) estimation was significant for all models (i.e., ranging between $\Delta\chi^2(4) = 9.81, p \leq 0.05$ for need satisfaction and $\Delta\chi^2(4) = 50.21, p \leq 0.001$ for need frustration). Indeed, with the exception of introjected regulation ($\chi^2 = 3.51, df = 1, p = 0.06$), differences between types of lessons (i.e., presence or absence of performance grading) were found for all variables, with students experiencing less intrinsic motivation ($\chi^2 = 43.07, df = 1, p \leq 0.001$), identified regulation ($\chi^2 = 13.91, df = 1, p \leq 0.001$) and need satisfaction ($\chi^2 = 4.91, df = 1, p \leq 0.05$), and more external regulation ($\chi^2 = 8.18, df = 1, p \leq 0.01$), amotivation ($\chi^2 = 43.43, df = 1, p \leq 0.001$), fear ($\chi^2 = 12.00, df = 1, p \leq 0.001$) and need frustration ($\chi^2 = 44.16, df = 1, p \leq 0.001$), during a lesson in which performance grading took place compared with a lesson in which no performance grading took place. Furthermore, these analyses served as a first step in the mediation analyses (i.e., second research question), because they give an indication of the total effect (τ) of the relation between performance grading and the motivational regulations and fear, without the inclusion of the mediators (see Tables 2 and 3).

In a second step, direct effects (τ') and indirect effects ($\alpha\beta$) were tested to observe if the significant associations between performance grading and motivational outcomes and fear were mediated by need satisfaction and need frustration. In the full model with need satisfaction as a mediator (see Table 3), a lowered effect size was found for the direct relationship between performance grading and intrinsic motivation (from $\tau = -0.31, p < 0.001$ to $\tau' = -0.26, p < 0.001$) and identified regulation (from $\tau = -0.17, p < 0.001$ to $\tau' = -0.11, p < 0.05$), indicating partial mediation ($\alpha\beta_{intrinsic} = 0.45, p \leq 0.001$ and $\alpha\beta_{identified} = 0.51, p \leq 0.001$). Because the relationship between performance grading and introjected regulation (from $\tau = 0.10, p = 0.06$ to $\tau' = 0.11, p \leq 0.05$), external regulation (from $\tau = 0.16, p \leq 0.01$ to $\tau' = 0.17, p \leq 0.01$), amotivation (from $\tau = 0.31, p \leq 0.001$ to $\tau' = 0.32, p \leq 0.001$) and fear (from $\tau = 0.17, p \leq 0.001$ to $\tau' = 0.18, p \leq 0.001$) did not significantly attenuate, need satisfaction was not considered as a mediator in these specific models (Cerin & MacKinnon, 2008).

In the full model with need frustration as a mediator (see Table 3), the direct relationship (τ') between performance grading and external regulation ($\tau' = -0.05, p = 0.36$) and fear ($\tau' = -0.06, p = 0.15$) was no longer significant, with need frustration fully mediating these relationships ($\alpha\beta_{external} = 0.62, p \leq 0.001$; $\alpha\beta_{fear} = 0.69, p \leq 0.001$). A lowered effect size was found for the direct relationship between performance grading and intrinsic motivation (from $\tau' = -0.31, p \leq 0.001$ to $\tau' = -0.26, p \leq 0.001$) and amotivation (from $\tau' = 0.31, p \leq 0.001$ to $\tau' = 0.16, p \leq 0.001$), indicating partial

Table 3
Students’ perceived intrinsic motivation, identified regulation, introjected regulation, external regulation, amotivation and level of fear in a lesson in which performance grading was applied compared with a lesson in which no performance grading took place, mediated by students’ perceived need satisfaction and need frustration.

	Model 2				Model 3							
	Total effect		Direct effect		Direct effect		Need frustration					
	τ	95% Cl _{bc}	τ'	95% Cl _{bc}	α -coefficient	95% Cl _{bc}	β -coefficient	95% Cl _{bc}	Specific indirect ($\alpha\beta$)	95% Cl _{bc}		
Quality of motivation												
Intrinsic motivation	-0.31***	(-0.40, -0.22)	-0.26***	(-0.35, -0.17)	0.47***	(0.41, 0.52)	0.45***	(0.39, 0.50)	0.33***	(0.28, 0.38)	-0.20***	(-0.22, -0.08)
Identified regulation	-0.17***	(-0.26, -0.08)	-0.11*	(-0.20, -0.02)	0.51***	(0.46, 0.57)	0.51***	(0.46, 0.57)	0.33***	(0.28, 0.38)	0.02	(0.05, 0.05)
Introjected regulation	0.10	(-0.00, 0.20)	0.11*	(0.01, 0.21)	0.14***	(0.08, 0.21)	0.15***	(0.08, 0.21)	0.33***	(0.28, 0.38)	-0.05, 0.09	(-0.02, 0.12)
External regulation	0.16**	(0.05, 0.26)	0.17**	(0.07, 0.28)	0.09**	(0.03, 0.16)	0.10*	(0.03, 0.17)	0.33***	(0.28, 0.38)	0.62	(0.56, 0.68)
Amotivation	0.31***	(0.22, 0.41)	0.32	(0.22, 0.41)	-0.12*	(-0.23, -0.01)	-0.03	(-0.07, 0.05)	0.33***	(0.28, 0.38)	0.51	(0.49, 0.51)
Fear	0.17***	(0.07, 0.27)	0.18***	(0.08, 0.27)	0.10**	(0.03, 0.16)	0.11*	(0.04, 0.17)	0.33***	(0.28, 0.38)	0.68	(0.63, 0.73)

Note. A positive relation was found between need satisfaction and need frustration (see supplementary online data), and an unexpected positive relationship was found between need satisfaction and external regulation and fear (see Table 3). However, when controlling for need frustration, the positive relationship between need satisfaction and external regulation ($\chi^2 = 0.44, df = 1, p = 0.51$) and fear ($\chi^2 = 2.50, df = 1, p = 0.11$) was no longer present, indicating that in the relationship between performance grading and negative motivational functioning, need frustration is the strongest clarifying variable.
* $p \leq 0.05$.
** $p \leq 0.01$.
*** $p \leq 0.001$.

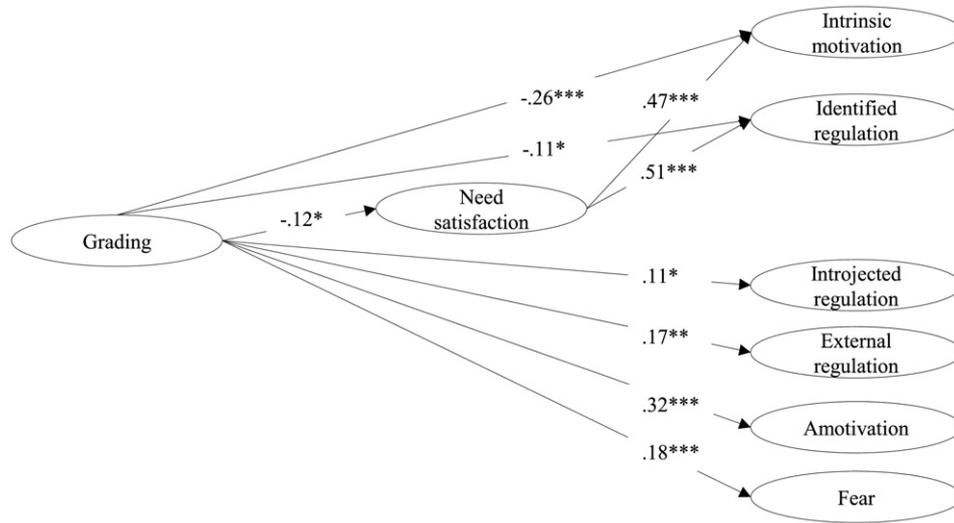


Fig. 1. Graphical representation of the direct relationships (τ), α and β coefficients as estimated in the full model with need satisfaction acting as a mediator. Note. β coefficients are only presented when need satisfaction was considered as a mediator (Cerin & MacKinnon, 2008). For introjected regulation, the direct relationship (τ) was found significant while the total effect (τ) was not found significant, indicating a suppressor effect. For all other variables, direct (τ) as well as total relationships (τ) were significant (see Table 3).

mediation ($\alpha\beta_{intrinsic} = -0.15, p \leq 0.001$ and $\alpha\beta_{amotivation} = 0.49, p \leq 0.001$). The full models with need satisfaction and need frustration proceeding as mediators are displayed graphically in respectively Fig. 1 and Fig. 2.

4. Discussion

Grounded in self-determination theory, the global purpose of the present study was to examine differences in students' motivational functioning, fear and need-based experiences between an authentic grading and non-grading PE class, and to examine the explanatory factors accounting for these differences. The context of the grading lesson in our study was a situation in which a multiple grading system was used in a highly 'visible' PE context (Trout & Graber, 2009). The awarded performance grade contributed to an average grade for PE, which is part of a yearly report. The average grades on this yearly report allow implicit and explicit ranking of students' performance, possibly triggering peer

comparison (Ames, 1992; Barenberg & Dutke, 2013; Elliot & Moller, 2003). In this context, the grading is most likely perceived to be evaluative and judgmental of one's performance.

4.1. Motivational differences and fear

Previous research (e.g., Butler, 1987; Butler & Nisan, 1986; Johnson et al., 2011; Pulfrey et al., 2011) has documented the motivational costs associated with performance grading. The present study replicates and extends this body of work by examining naturally occurring motivational and fear-related differences in real-life grading and non-grading PE lessons. Also, while previous studies have looked into composite scores of autonomous and controlled motivation (e.g., Aelterman et al., 2012; Mouratidis, Vansteenkiste, Lens, & Sideridis, 2008; Pulfrey et al., 2011), herein, we have examined in greater detail whether different subtypes of both autonomous and controlled motivation vary as a function of being exposed to a grading and non-grading

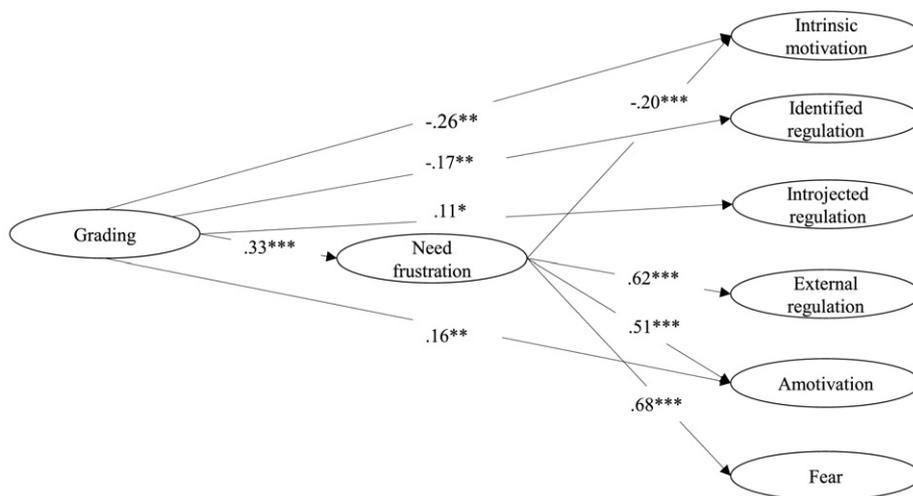


Fig. 2. Graphical representation of the direct relationships (τ), α and β coefficients as estimated in the full model with need frustration acting as a mediator. Note. β coefficients are only presented when need frustration was considered as a mediator (Cerin & MacKinnon, 2008). For introjected regulation, the direct relationship (τ) was found significant while the total effect (τ) was not found significant, indicating a suppressor effect. For intrinsic motivation, identified regulation and amotivation, direct (τ) as well as total relationships (τ) were found significant (see Table 3).

class. The necessity to look at subtypes, as advocated by some scholars in the field (Gagné et al., 2014; Taylor et al., 2014), was supported in the current study as not all forms of controlled regulation varied in parallel.

Specifically, during PE lessons in which performance grading occurs, students find the lesson to be less interesting and enjoyable (i.e., intrinsic motivation) and perceived the lesson to be less meaningful (i.e., identified regulation). These findings are in line with previous studies, which showed that performance-based grading undermines love of learning, interest and curiosity (i.e., less intrinsic motivation; Butler & Nisan, 1986; Butler, 1987; Grolnick & Ryan, 1987; Johnson et al., 2011; Pulfrey et al., 2011, 2013) and undermines the relevance of participating in the PE lesson (Johnson et al., 2011; Pulfrey et al., 2011). The costs associated with performance grading also manifested through the presence of more maladaptive forms of motivation. That is, during performance grading classes, students reported more external regulation, being more amotivated and experienced more fear. These findings are consistent with those from two other studies, in which it was found that students in a graded condition (i.e., students who were judged on their performance by means of a grade) experienced more pressure (i.e., external regulation Grolnick & Ryan, 1987), and girls who participated in norm-referenced PE assessments experienced more external regulation and amotivation (Johnson et al., 2011). In addition, in previous work, students reported also more negative emotional reactions such as fear when exposed to performance grading (McDonald, 2001; Putwain & Best, 2011).

Yet, in the present study, no differences in introjected regulation emerged as a function of grading. Thus, whereas the pressure imposed by someone else (i.e., external regulation) augmented as a function of performance grading, the pressure imposed by one's self (i.e., introjected regulation) did not increase. This is an interesting and somewhat unexpected finding by itself because one may expect that under grading circumstances, students become increasingly concerned with their self-worth and consider the graded activity as a means to impress others. Future studies may examine whether such internal pressures get activated under particular circumstances.

4.2. Explanatory mechanisms: need-based experiences

The second important aim of the present study was to examine whether need-based experiences would account for any observed motivational differences between grading and non-grading lessons. Following recent developments, we considered both the role of the satisfaction and frustration of students' psychological needs for autonomy, competence, and relatedness (Vansteenkiste & Ryan, 2013). The inclusion of both constructs was critical as need satisfaction predominantly accounted for the link between performance grading and the more self-determined forms of motivation, while need frustration largely explained the less self-determined motivational outcomes and fear. That is, when students were exposed to performance grading, they experienced a lack of choice or freedom (i.e., low autonomy satisfaction), a sense of disconnection to others (i.e., low relatedness satisfaction) and a sense of not being able to reach the criteria (i.e., low competence satisfaction), which then led students to find the lesson less enjoyable (i.e., intrinsic motivation) and valuable (i.e., identified motivation).

Furthermore, as a function of performance grading, students not only reported less need satisfaction, they also felt more pressured to perform well (i.e., high autonomy frustration), were more likely to feel rejected by others (i.e., high relatedness frustration) and more strongly felt like a failure (i.e., high competence frustration). In a similar vein as low need satisfaction partially explained why students found the grading lesson less enjoyable, also experienced need frustration partially explained why students experienced less joy as a function of performance grading. While previous studies already showed that lower levels of experienced autonomy satisfaction and competence satisfaction explained the relation between performance grading and intrinsic

motivation (Pulfrey et al., 2013), the current study adds to this literature by showing that this relationship is also partially explained by feelings of need frustration.

Moreover, students who reported higher levels of need frustration because of being exposed to performance grading, were more likely to put effort into the lesson out of external pressure (i.e., external regulation), were more likely to lack motivation (i.e., amotivation), or to experience fear. Experiences of need frustration, rather than need satisfaction thus explained differences in external regulation, amotivation and fear (also see Bartholomew et al., 2011a; Haerens et al., 2015).

Interestingly, students' perceived need satisfaction was positively correlated with students' perceived need frustration (see supplementary online data). However, this positive link was only shown during lessons in which performance grading took place. It might have been the case, that students experienced alternating episodes of need satisfaction and need frustration during performance grading. For instance, in the beginning of the lesson, a student might be uncertain about the quality of his performance (i.e., low need satisfaction) and might feel pressured to perform well (i.e., high need frustration). Yet, after performing well during the grading activity, the student might feel capable of his functioning (i.e., high need satisfaction) and the pressure might fade (i.e., low need frustration). Because such episodes of need satisfaction and need frustration were aggregated throughout the entire lesson, these dynamics might possibly explain why we found a positive relation between need satisfaction and need frustration.

4.3. Strengths, limitations and directions for future research

One strength of this study was the use of multi-level regression analyses and more specifically its evaluation of variances at the class, student and measurement level. These analyses revealed that for all motivational outcomes and fear, variances were significantly different from zero at all levels. This suggests that there might be class level factors (e.g., the way the lesson is taught, the way students were graded, the objectives of the lesson) as well as student level factors (e.g., overall motivation for physical education) that can explain motivational differences. Further, these analyses suggested that students experienced a substantial amount of variation in their motivational functioning from lesson to lesson. This implies that there might be within-student level factors (e.g., the provided extent of individualised feedback in both lessons) that explain differences in motivational functioning from lesson to lesson. Since these differences are substantial, this topic could be interesting to explore in future research.

The fact that this study was purposefully situated in the PE context might be regarded as another strength. It was implied that, due to high 'visibility' of performances, this particular context might make students' experiences even more salient (Trout & Graber, 2009). However, that does not imply that the presence of grading in a more academic environment (e.g., maths or literature) might not come at a motivational and affective price. General aspects (e.g., criterion referenced grading and judgment of product performance) of the multiple grading system presented in this study, are existing in other, more academic contexts as well. Given these common grounds, the results found in this study may potentially be generalised to more academic settings (Barenberg & Dutke, 2013; Butler, 1987; Butler & Nisan, 1986; Pulfrey et al., 2011, 2013), an issue that deserves further research.

The present study also has several limitations. First, because students' skill level may have developed over time, this could have interfered with students' feelings of competence (or self-concept about their abilities) which would then possibly reduce the negative impact of performance grading. Several strategies could have been used to control for this issue. For instance, we could have (a) counterbalanced the design with non-grading lessons following grading lessons in half of the classes, (b) measured students' skill level in both lessons as to include it as a covariate, and (c) included a control group that was not performance graded. Yet, this was not attainable in the current study given

that it was conducted in a real-life, ecologically valid setting. Also, given that the time frame between both measurements was in most classes one to three weeks, we consider the learning effects to be only of minimal influence.

Second, in the present research we chose a within-student design in which students were measured after a non-grading lesson and after a grading lesson. To get a more refined understanding of students' motivational and affective experiences in relation to performance grading, it would have been even stronger if we had also measured students just before the performance grading lessons, thereby tapping into their anticipated motivation, fear and need-based experiences for the upcoming lesson.

Third, although video observations gave insight in the performance-based assessment practices, these observations did not allow us to provide insight whether the assessment criteria were aligned with the content standards (e.g., practicing basketball techniques and being assessed on those techniques versus being assessed on playing a match; Biggs, 1996; Borghouts et al., 2016).

Fourth, no basis was provided to suggest that the presence of performance grading is highly important when explaining students' motivational functioning in PE. It was interesting to note that, although results indicated statistically significant differences in students' motivation, fear and need-based experiences, the differences between both lessons were rather small (i.e., small effect sizes) and the variable 'grading lesson' only explained small amounts of variance. Other factors such as whether the teacher provides insight in assessment criteria or gives feedback (Sadler, 1989), or differences in teachers' motivating style between both lessons (De Meyer et al., 2014; Reeve, Jang, Carrell, Jeon, & Barch, 2004) that go well beyond the mere presence of a performance grade, might potentially be of greater influence (Ryan & Brown, 2005; Ryan & Weinstein, 2009). It warrants further investigation as to whether it is the presence of performance grading in itself, the lesson content, or that the way the lessons are taught with teachers possibly taking up a more controlling stance when grading, that explain students' motivation.

Fifth, it remains unclear if these negative outcomes represent incidental or lasting experiences and if these negative outcomes affect students' learning in PE. Therefore, it is recommended that future research develops a longitudinal design in which students' motivational functioning, fear and need-based experiences are followed over a greater period of time and in different domains of sports. As such, more detailed insights into students' motivational functioning, fear and need-based experiences may be yielded, when being graded in different sports. Also learning progress could be included as an outcome.

4.4. Implications for education

Results from the present study were gathered in an educational environment in which students were awarded grades that served as a judgment of their performance. Findings suggest that it is important for teachers to reflect on the meaning or functional significance grading has for students in their educational practice. However, a critical reflection on the curriculum is not only the teachers' responsibility. The extent to which teachers grade their students is also partly due to reasons of selection (Newton, 2007). Thus, in pursuit of positive motivational and affective experiences, we argue that this responsibility should be shared with school boards and policy-makers (Yu, Chen, Levesque-Bristol, & Vansteenkiste, 2016).

Whilst students are subjected to performance grades, it seems important for teachers to induce feelings of choice or freedom, feelings of connection to others and opportunities to reach criteria (i.e., need satisfaction) and to reduce feelings of pressure to perform well, feelings of rejection by others and feelings of failure (i.e., need frustration), in order to stimulate positive motivational and affective experiences. This does not imply that, from a motivational perspective, it is per definition unfavourable to apply grading in education. There might be

conditions under which grading does not consist a need undermining or frustrating event and may even be conducive to students' needs as well as most volitional forms of motivation (Maes et al., in preparation).

From an SDT point of view and attempting to stimulate students' needs and most volitional forms of motivation in education, assessment with the aim of grading can be applied with an informational function (Ryan & Weinstein, 2009). An informational assessment is referred to when teachers deploy assessment as a non-controlling means to improve learning. When grading students, it is important to follow up with means to improve learning by using strategies that go beyond providing grades, such as providing transparent criteria, discussing assessments among each other, actively involving students within the learning process and providing insight in subsequent learning objectives (Pat-El & Van der Poel, 2011). Thus, the issue raised in the present research is not merely related to the presence of grading in itself, but to what extent assessment is used solely with the function of judging performance rather than with a focus on learning. Unravelling the relation between different functions and meanings of assessment and its motivational outcomes is something that merits further investigation.

4.5. Conclusion

This study provides further insight in students' motivational and affective experiences as a function of performance grading. Existing literature has already shown that performance grading potentially undermines more volitional forms of motivation (e.g., Butler & Nisan, 1986; Johnson et al., 2011; Pulfrey et al., 2011) and that this relationship can potentially be explained by lowered experiences of autonomy and competence satisfaction (Pulfrey et al., 2013). The present study adds to this literature by highlighting that the relation between performance grading and intrinsic motivation, as well as negative motivational functioning, can be explained by increased feelings of need frustration. From a practical point of view, since performance grading is part of PE assessment, it seems important for teachers to carefully reflect on their curricula and their current way of assessing, particularly within a highly 'visible' educational environment, where positive motivational and affective experiences are pursued.

Funding

This work was supported by The Netherlands Organisation for Scientific Research (Grant 023.004.015).

Appendix A. Supplementary data

Supplementary data to this article can be found online at <http://dx.doi.org/10.1016/j.lindif.2017.03.017>.

References

- Aelterman, N., Vansteenkiste, M., Van Keer, H., van den Berghe, L., de Meyer, J., & Haerens, L. (2012). Students' objectively measured physical activity levels and engagement as a function of between-class and between-student differences in motivation toward physical education. *Journal of Sport & Exercise Psychology, 34*(4), 457–480.
- Ames, C. (1992). Classrooms: Goals, structures, and student motivation. *Journal of Educational Psychology, 84*(3), 261–271.
- Amrein, A. L., & Berliner, D. (2002). High-stakes testing, uncertainty, and student learning. *Education Policy Analysis Archives, 10*(18), 1–50.
- Annerstedt, C., & Larsson, S. (2010). 'I have my own picture of what the demands are...': Grading in Swedish PEH-problems of validity, comparability and fairness. *European Physical Education Review, 16*(2), 97–115. <http://dx.doi.org/10.1177/1356336X10381299>.
- Barenberg, J., & Dutke, S. (2013). Metacognitive monitoring in university classes: Anticipating a graded vs. a pass-fail test affects monitoring accuracy. *Metacognition and Learning, 8*(2), 121–143. <http://dx.doi.org/10.1007/s11409-013-9098-3>.
- Bartholomew, K. J., Ntoumanis, N., Ryan, R. M., Bosch, J. A., & Thøgersen-Ntoumani, C. (2011a). Self-determination theory and diminished functioning: The role of interpersonal control and psychological need thwarting. *Personality and Social Psychology Bulletin, 37*(11), 1459–1473. <http://dx.doi.org/10.1177/0146167211413125>.

- Bartholomew, K. J., Ntoumanis, N., Ryan, R. M., & Thøgersen-Ntoumani, C. (2011b). Psychological need thwarting in the sport context: Assessing the darker side of athletic experience. *Journal of Sport & Exercise Psychology*, 33(1), 75–102.
- Biggs, J. (1996). Enhancing teaching through constructive alignment. *Higher Education*, 32, 347–364.
- Borghouts, L., Slingerland, M., & Haerens, L. (2016). Assessment quality and practices in secondary PE in the Netherlands. *Physical Education and Sport Pedagogy*, 1–17. <http://dx.doi.org/10.1080/17408989.2016.1241226>.
- Butler, R. (1987). Task-involving and ego-involving properties of evaluation: Effects of different feedback conditions on motivational perceptions, interest, and performance. *Journal of Educational Psychology*, 79(4), 474–482.
- Butler, R., & Nisan, M. (1986). Effects of no feedback, task-related comments, and grades on intrinsic motivation and performance. *Journal of Educational Psychology*, 78(3), 210–216.
- Cerin, E., & MacKinnon, D. P. (2008). A commentary on current practice in mediating variable analyses in behavioural nutrition and physical activity. *Public Health Nutrition*, 12(8), 1182–1188. <http://dx.doi.org/10.1017/S1368980008003649>.
- Chan, K., Hay, P., & Tinning, R. (2011). Understanding the pedagogic discourse. *Asia-Pacific Journal of Health, Sport and Physical Education*, 2(1), 3–18.
- Chen, B., Vansteenkiste, M., Beyers, W., Boone, L., Deci, E. L., Van der Kaap-Deeder, J., ... Verstuyf, J. (2015). Basic psychological need satisfaction, need frustration, and need strength across four cultures. *Motivation and Emotion*, 39(2), 216–236. <http://dx.doi.org/10.1007/s11031-014-9450-1>.
- De Meyer, J., Tallir, I. B., Soenens, B., Vansteenkiste, M., Aelterman, N., Van Den Berghe, L., ... Haerens, L. (2014). Does observed controlling teaching behavior relate to students' motivation in physical education? *Journal of Educational Psychology*, 106(2), 541–554. <http://dx.doi.org/10.1037/a0034399>.
- Deci, E. L., & Ryan, R. M. (2000). The 'what' and 'why' of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry*, 11(4), 227–268. http://dx.doi.org/10.1207/S15327965PLI1104_01.
- Elliot, A. J., & McGregor, H. A. (1999). Test anxiety and the hierarchical model of approach and avoidance achievement motivation. *Journal of Personal and Social Psychology*, 76(4), 628–644.
- Elliot, A. J., & Moller, A. C. (2003). Performance-approach goals: Good or bad forms of regulation? *International Journal of Educational Research*, 39(4–5), 339–356. <http://dx.doi.org/10.1016/j.ijer.2004.06.003>.
- European Commission/EACEA/Eurydice (2013). *Physical education and sport at school in Europe*. Luxembourg: Eurydice Report. <http://dx.doi.org/10.2797/49648>.
- Gagné, M., Forest, J., Vansteenkiste, M., Crevier-Braud, L., van den Broeck, A., Aspli, A. K., ... Westbye, C. (2014). The multidimensional work motivation scale: Validation evidence in seven languages and nine countries. *European Journal of Work and Organizational Psychology*, 0(0), 1–19. <http://dx.doi.org/10.1080/1359432X.2013.877892>.
- Grolnick, W. S., & Ryan, R. M. (1987). Autonomy in children's learning: An experimental and individual difference investigation. *Journal of Personality And Social Psychology*, 52(5), 890–898. <http://dx.doi.org/10.1037/0022-3514.52.5.890>.
- Guay, F., Chantal, J., Ratelle, C. F., Marsh, H. W., Larose, S., & Boivin, M. (2010). Intrinsic, identified, and controlled types of motivation for school subjects in young elementary school children. *The British Journal of Educational Psychology*, 80, 711–735. <http://dx.doi.org/10.1348/000709910X499084>.
- Haerens, L., Aelterman, N., Vansteenkiste, M., Soenens, B., & van Petegem, S. (2015). Do perceived autonomy-supportive and controlling teaching relate to physical education students' motivational experiences through unique pathways? Distinguishing between the bright and dark side of motivation. *Psychology of Sport and Exercise*, 16, 26–36. <http://dx.doi.org/10.1016/j.psychsport.2014.08.013>.
- Hay, P. J., & Macdonald, D. (2008). (Mis)appropriations of criteria and standards-referenced assessment in a performance-based subject. *Assessment in Education: Principles, Policy & Practice*, 15(2), 153–168. <http://dx.doi.org/10.1080/09695940802164184>.
- Hox, J. (2010). *Multilevel analysis: Techniques and applications* (2nd ed.). New York, NY: Routledge.
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6(1), 1–55. <http://dx.doi.org/10.1080/10705519909540118>.
- Jang, H., Kim, E. J., & Reeve, J. (2016). Why students become more engaged or more disengaged during the semester: A self-determination theory dual-process model. *Learning and Instruction*, 43, 27–38. <http://dx.doi.org/10.1016/j.learninstruc.2016.01.002>.
- Johnson, T. G., Prusak, K. A., & Pennington, T. (2011). The effects of the type of skill test, choice, and gender on the situational motivation of physical education students. *Journal of Teaching in the Physical Education*, 30(3), 281–295.
- Jones, B. D. (2007). The Unintended outcomes of high stakes testing. *Journal of Applied School Psychology*, 23(2), 65–86. http://dx.doi.org/10.1300/J370v23n02_05.
- Kline, R. (2011). *Principles and practice of structural equation modeling* (3rd ed.). New York: The Guilford Press.
- Lingard, B. (2010). Policy borrowing, policy learning: Testing times in Australian schooling. *Critical Studies in Education*, 51(2), 129–147. <http://dx.doi.org/10.1080/17508481003731026>.
- López-Pastor, V. M., Kirk, D., Lorente-Catalán, E., MacPhail, A., & Macdonald, D. (2012). Alternative assessment in physical education: A review of international literature. *Sport, Education and Society*, 18(1), 57–76. <http://dx.doi.org/10.1080/13573322.2012.713860>.
- Maes, J., Krijgsman, C., Cardon, G., Tallir, I., Borghouts, L., & Haerens, L. (Unpublished manuscript). How does assessment for learning in physical education relate to students' motivation, perceived competence and level of fear? Manuscript in preparation.
- McDonald, A. S. (2001). The prevalence and effects of test anxiety in school children. *Educational Psychology*, 21(1), 89–101. <http://dx.doi.org/10.1080/01443410124641>.
- 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 & Exercise Psychology*, 30(2), 240–268.
- Muthén, L. K., & Muthén, B. O. (2015). *Mplus user's guide* (7th ed.). Los Angeles, CA: Muthén & Muthén.
- Newton, P. E. (2007). Clarifying the purposes of educational assessment. *Assessment in Education*, 14(2), 149–170. <http://dx.doi.org/10.1080/09695940701478321>.
- Niemiec, C. P., & Ryan, R. M. (2009). Autonomy, competence, and relatedness in the classroom: Applying self-determination theory to educational practice. *Theory and Research in Education*, 7(2), 133–144. <http://dx.doi.org/10.1177/1477878509104318>.
- Ntoumanis, N. (2001). A self-determination approach to the understanding of motivation in physical education. *British Journal of Educational Psychology*, 71(2), 225–242. <http://dx.doi.org/10.1348/000709901158497>.
- Ntoumanis, N., & Standage, M. (2009). Motivation in physical education classes: A self-determination theory perspective. *Theory and Research in Education*, 7(2), 194–202. <http://dx.doi.org/10.1177/1477878509104324>.
- Pat-El, R., & van der Poel, M. (2011). Opmattingen van leerkrachten over evalueren om te leren. In J. Castelijn, M. Segers, & K. Struyven (Eds.), *Evalueren om te leren* (pp. 181–191). Bussum: Coutinho.
- Pulfrey, C., Buchs, C., & Butera, F. (2011). Why grades engender performance-avoidance goals: The mediating role of autonomous motivation. *Journal of Educational Psychology*, 103(3), 683–700. <http://dx.doi.org/10.1037/a0023911>.
- Pulfrey, C., Darnon, C., & Butera, F. (2013). Autonomy and task performance: Explaining the impact of grades on intrinsic motivation. *Journal of Educational Psychology*, 105(1), 39–57. <http://dx.doi.org/10.1037/a0029376>.
- Putwain, D. W., & Best, N. (2011). Fear appeals in the primary classroom: Effects on test anxiety and test grade. *Learning and Individual Differences*, 21, 580–584. <http://dx.doi.org/10.1016/j.lindif.2011.07.007>.
- Rasbash, J., Steele, F., Browne, W., & Goldstein, H. (2014). *A user's guide to MLwiN, v2.31*. Centre for Multilevel Modelling, University of Bristol.
- Redelius, K., & Hay, P. J. (2012). Student views on criterion-referenced assessment and grading in Swedish physical education. *Physical Education and Sport Pedagogy*, 17(2), 211–225. <http://dx.doi.org/10.1080/17408989.2010.548064>.
- Reeve, J., Jang, H., Carrell, D., Jeon, S., & Barch, J. (2004). Enhancing students' engagement by increasing teachers' autonomy support. *Motivation and Emotion*, 28(2), 147–169.
- Ryan, R. M., & Brown, K. W. (2005). Legislating competence: High-stakes testing policies and their relations with psychological theories and research. In A. J. Elliot, & C. S. Dweck (Eds.), *Handbook of competence and motivation* (pp. 354–372). London: The Guilford Press.
- Ryan, R. M., & Weinstein, N. (2009). Undermining quality teaching and learning: A self-determination theory perspective on high-stakes testing. *Theory and Research in Education*, 7(2), 224–233. <http://dx.doi.org/10.1177/1477878509104327>.
- Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, 18, 119–144.
- Schaffner, E., & Schiefele, U. (2007). The effect of experimental manipulation of student motivation on the situational representation of text. *Learning and Instruction*, 17, 755–772. <http://dx.doi.org/10.1016/j.learninstruc.2007.09.015>.
- Strain, M. (2009). The education reform act 1988: Success or failure? A short report of a BELMAS discussion forum. *Management in Education*, 23(4), 151–154. <http://dx.doi.org/10.1177/0892020609344036>.
- Svennberg, L., Meckbach, J., & Redelius, K. (2014). Exploring PE teachers' 'gut feelings': An attempt to verbalise and discuss teachers' internalised grading criteria. *European Physical Education Review*, 20(2), 199–214. <http://dx.doi.org/10.1177/1356336X13517437>.
- Taylor, G., Jungert, T., Mageau, G. A., Schattke, K., Dedic, H., Rosenfield, S., & Koestner, R. (2014). A self-determination theory approach to predicting school achievement over time: The unique role of intrinsic motivation. *Contemporary Educational Psychology*, 39(4), 342–358. <http://dx.doi.org/10.1016/j.cedpsych.2014.08.002>.
- Trout, J., & Graber, K. C. (2009). Perceptions of overweight students concerning their experiences in physical education. *Journal of Teaching in Physical Education*, 28, 272–292.
- Tunstall, P., & Gipps, C. (1996). Teacher feedback to young children in formative assessment: A typology. *British Educational Research Journal*, 22(4), 389–404.
- van der Kaap-Deeder, J., Soenens, B., Boone, L., Vandenkerckhove, B., Stembée, E., & Vansteenkiste, M. (2016). Evaluative concerns perfectionism and coping with failure: Effects on rumination, avoidance, and acceptance. *Personality and Individual Differences*, 101, 114–119. <http://dx.doi.org/10.1016/j.paid.2016.05.063>.
- Vansteenkiste, M., & Ryan, R. M. (2013). On psychological growth and vulnerability: Basic psychological need satisfaction and need frustration as a unifying principle. *Journal of Psychotherapy Integration*, 23(3), 263–280. <http://dx.doi.org/10.1037/a0032359>.
- Vansteenkiste, M., Simons, J., Lens, W., Sheldon, K. M., & Deci, E. L. (2004). Motivating learning, performance, and persistence: The synergistic effects of intrinsic goal contents and autonomy-supportive contexts. *Journal of Personality And Social Psychology*, 87(2), 246–260. <http://dx.doi.org/10.1037/0022-3514.87.2.246>.
- Vansteenkiste, M., Lens, W., & Deci, E. L. (2006). Intrinsic versus extrinsic goal contents in self-determination theory: Another look at the quality of academic motivation. *Educational Psychologist*, 41(1), 19–31. http://dx.doi.org/10.1207/s15326985ep4101_4.
- Vansteenkiste, M., Ryan, R. M., & Deci, E. L. (2008). Self-determination theory and the explanatory role of psychological needs in human well-being. In L. Bruni, F. Comim, & M. Pugno (Eds.), *Capabilities and happiness* (pp. 187–223). Oxford, UK: Oxford University Press. <http://dx.doi.org/10.1017/S0266267111000058>.
- Weinstein, C. (1987). *LASSI user's manual*. Clearwater, FL: H&H Publishing.
- Yu, S., Chen, B., Levesque-Bristol, C., & Vansteenkiste, M. (2016). Chinese education examined via the lens of self-determination. *Educational Psychology Review*, 1–38. <http://dx.doi.org/10.1007/s10648-016-9395-x>.