



ELSEVIER

Contents lists available at ScienceDirect

Journal of School Psychology

journal homepage: www.elsevier.com/locate/jschpsyc

Does teacher evaluation based on student performance predict motivation, well-being, and ill-being?

Ricardo Cuevas^a, Nikos Ntoumanis^b, Juan G. Fernandez-Bustos^{a,*},
Kimberley Bartholomew^c

^a Faculty of Education, University of Castilla-La Mancha, Spain

^b School of Psychology & Speech Pathology, Curtin University, Australia

^c School of Education and Lifelong Learning, University of East Anglia, United Kingdom

ARTICLE INFO

Action Editor: Amy Briesch

Keywords:

Self-determination

Vitality

Burnout

Mediation

Structural equation model

ABSTRACT

This study tests an explanatory model based on self-determination theory, which posits that pressure experienced by teachers when they are evaluated based on their students' academic performance will differentially predict teacher adaptive and maladaptive motivation, well-being, and ill-being. A total of 360 Spanish physical education teachers completed a multi-scale inventory. We found support for a structural equation model that showed that perceived pressure predicted teacher autonomous motivation negatively, predicted amotivation positively, and was unrelated to controlled motivation. In addition, autonomous motivation predicted vitality positively and exhaustion negatively, whereas controlled motivation and amotivation predicted vitality negatively and exhaustion positively. Amotivation significantly mediated the relation between pressure and vitality and between pressure and exhaustion. The results underline the potential negative impact of pressure felt by teachers due to this type of evaluation on teacher motivation and psychological health.

1. Introduction

The Teaching and Learning International Survey (TALIS) of the Organization for Economic Co-operation and Development (OECD, 2014) notes that the most important factor for quality education is teacher performance. Teachers are the main agents for engaging students in school tasks and promoting their learning (Rockoff, 2004). In this regard, numerous studies have examined ways in which to measure teacher performance (e.g., Pas, Bradshaw, & Hershfeldt, 2012). Traditionally, there are two elements involved in evaluating teacher performance: supervision (the formative aspect) and evaluation (the summative aspect). Supervision involves the assessment of teachers' lesson plans, their teaching skills and instructional strategies, and how well they have mastered the material. On the other hand, the summative element involves evaluating how well the students have learned the lesson content delivered by the teacher based on the students' performance on assessments or their grades.

In this paper, we focus on teacher evaluation that is based on student performance. According to the TALIS report (Ministry of Education, Culture and Sport, 2014), the most widely used procedure for teacher evaluation across several countries (e.g., United Kingdom, Sweden, France) of the Organization for Economic Co-operation and Development (OECD) is based on students' academic grades. In Spain, for example, where current legislation stipulates the need to evaluate the performance of teachers (Marina, Pellicer, & Manso, 2015), the most widely used procedure for teacher evaluation (used in 97% of schools) is based on students' academic

* Corresponding author at: Faculty of Education, University of Castilla-La Mancha, Plaza de la Universidad, 3, 02071 Albacete, Spain.

E-mail addresses: ricardo.cuevas@uclm.es (R. Cuevas), juang.fernandez@uclm.es (J.G. Fernandez-Bustos).

<https://doi.org/10.1016/j.jsp.2018.03.005>

Received 27 May 2016; Received in revised form 26 December 2017; Accepted 12 March 2018.

0022-0815/© 2018 Society for the Study of School Psychology. Published by Elsevier Ltd. All rights reserved.

grades. The respective “weight” or importance of students' grades in the evaluation of teachers' performance can, however, vary from one school to another and students' grades may also be used alongside other forms of evaluation in Spain [e.g., classroom observation (59% of schools), student surveys (72%), or knowledge assessment (34%) (OECD, 2014)]. Principals who conduct formal evaluations of their teachers reported that their evaluations could affect career progress, changes in work responsibilities or, in some extreme cases, the dismissal of teachers (OECD, 2014). Perhaps not surprisingly, most of the teachers questioned in the TALIS reported disagreement with the current evaluation system and with the feedback they receive through this process (Ministry of Education, Culture and Sport, 2014). Given that the use of teacher evaluation based on students' performance has spread considerably in various countries in recent years (Isore, 2009), research on the impacts of this type of evaluation is important.

Despite its increasing use in various countries, some authors have suggested that external incentives are not always effective in improving teaching performance. For example, Yuan et al. (2012) found that incentive pay programs did not improve the practices and motivation of teachers. Furthermore, in a review article, Firestone (2014) indicated that incentives programs that use performance-based pay in educational contexts to improve student performance are ineffective and can undermine the intrinsic incentives of the teachers.

The practice of providing external incentives to teachers that are contingent on their students' performance is linked to growth models for evaluation in education. Such models aspire to measure the specific contribution of teachers to the growth of their students (McCaffrey, Lockwood, Koretz, & Hamilton, 2003). Value-added models are some of the most widely used types of growth models. Specifically, these models try to capture student performance over time (i.e., the development of knowledge or skills) as a consequence of student experiences in schools (Harvey, 2004). Despite their widespread use, the effectiveness of value-added models has been questioned on a number of grounds. For example, research has shown that school factors (including teacher performance) account for only about 20% of the variance in student performance (Berliner, 2014). In addition, Rothstein (2010) identified several additional external factors that can influence student performance, including students' experiences with previous teachers, the simultaneous influence of different teachers, the number of students in the class, the inclusion of students with special educational needs, curriculum materials, and the sociodemographic characteristics of the school and its students.

Given the number of different factors that may interact to influence student performance — many of which are outside of the teacher's control — it has been argued that student performance outcomes do not appear to be a sufficiently robust means by which to assess teacher effectiveness or to warrant consideration in decisions that may affect a teacher's career (McCaffrey, Sass, Lockwood, & Kata, 2009). In fact, various negative consequences resulting from the use of this type of evaluation have recently been documented, including a compression of the curriculum, decreased collaborative work between educators, and discouragement of teachers to work with the neediest students (Baker et al., 2010; Hewitt, 2015). In addition, there is growing evidence to indicate that the pressure felt by teachers as a result of this form of evaluation may also have negative repercussions for their psychological health.

1.1. Teacher evaluation and psychological health

The limited research examining the impact of teacher evaluation based on student performance has shown that teachers experience increased stress, pressure, and anxiety as a consequence of such evaluations (Hewitt, 2015; Jiang, Spörte, & Lupescu, 2015; von der Embse, Pendergast, Segool, Saeki, & Ryan, 2016). For example, Goldhaber and Hannaway (2004) found that evaluation-related pressure and anxiety levels were high not only among teachers in the U.S. whose schools had poor results and who attempted to improve those results, but also among teachers whose schools exhibited high performance and who tried to maintain that high level. Furthermore, accountability for student outcomes has been shown to be associated with increased teacher anxiety, and decreased teacher motivation, particularly among teachers who do not achieve the objectives set by the administration, irrespective of how much they have endeavored to do so (Feng, Figlio, & Sass, 2010; Finnigan & Gross, 2007).

Despite increasing interest in the topic, the specific impact of student performance-based teacher evaluation on the well-being and ill-being of teachers needs more empirical attention in an effort to explore mechanisms that could mediate such an impact. (Taylor & Tyler, 2012). To this end, Self-Determination Theory (SDT; Deci & Ryan, 1985) and its focus on motivation- may prove as a helpful conceptual framework.

1.2. Self-determination theory

Various studies (e.g., Cuevas, Sanchez-Oliva, Bartholomew, Ntoumanis, & Garcia-Calvo, 2015; Taylor & Ntoumanis, 2007; Taylor, Ntoumanis, & Standage, 2008) have noted the usefulness of Self-Determination Theory (SDT; Deci & Ryan, 1985) for the study of teacher motivation and psychological health. SDT is a widely applied theoretical approach to the study of human motivation, development, and well-being. The theory focuses on different “types” of motivation which have been shown to predict a diverse range of adaptive and maladaptive cognitive, affective, and behavioral outcomes. Specifically, Deci and Ryan (1985) differentiated between three different forms of motivation. First, autonomous motivation is volitional and reflects interest or personal value. For example, when an activity is performed for pleasure or personal growth. Second, controlled motivation reflects external and/or internal contingencies and pressures. For example, when an activity is performed for external incentives, such as money or social recognition. Finally, amotivation reflects a lack of both intrinsic and extrinsic motivation. Individuals engage passively in activities without any sense of intention. Several studies have linked teacher motivation with teacher well-being or ill-being. For instance, higher scores on autonomous motivation have been shown to be positively associated with higher levels of well-being, and negatively associated with higher levels of ill-being in Israeli teachers (Roth, Assor, Kanat-Maymon, & Kaplan, 2007). In the same country, Eyal and Roth (2011) found that burnout in teachers was negatively predicted by autonomous motivation and positively predicted by controlled

motivation. In addition, [Fernet, Guay, Senécal, and Austin \(2012\)](#) found that autonomous motivation negatively predicted emotional exhaustion in French-Canadian teachers.

Although the association between teacher motivation and well-being/ill-being has been explored within the literature, the influence of teacher evaluations on both adaptive and maladaptive types of teacher motivation and their psychological health is less well understood. Given that a lack of personal control has been associated with ill-being ([Weiner, 2004](#)), it may be particularly important to consider whether the link between student performance-contingent evaluations and teachers' psychological health is related to the perceived lack of control, or self-determination, that teachers perceive in relation to their students' performance ([Berliner, 2014](#); [Konstantopoulos, 2014](#)). Specifically, evaluations based on external and largely non-controllable criteria, such as student performance, are likely to be perceived as controlling and hence have the potential to undermine self-determined motivation ([Deci & Ryan, 1985](#)). In such situations, teachers are likely to feel controlled in their motivation to work or even demotivated. In turn, such motivational states are unlikely to be conducive to the nurturing of one's well-being. Despite this, the role of teacher motivation as a mediator between student performance-contingent teacher evaluation and well-being and ill-being has not been explored.

1.3. The present study

In this study, we propose and test a model, using data collected from physical education teachers, that links perceived pressure due to evaluations dependent on student performance, with teacher motivation and, in turn, well-being and ill-being. Whilst we acknowledge that teacher evaluation consists of multiple dimensions ([Isore, 2009](#)), we are specifically interested in this particular type of evaluation because it is becoming increasingly used in educational settings and has the potential to undermine the psychological health of teachers. Specifically, the objectives of the study were to (a) analyze the association between the perceived pressure caused by teacher evaluation based on student performance and teacher psychological well-being and ill-being and (b) examine the different types of teacher motivation, as outlined by SDT, as possible mediators of this association. To this end, four hypotheses were proposed. First, based on previous research that has indicated that teacher evaluations influence teacher well-being and ill-being (e.g., [Dworkin & Tobe, 2014](#); [Hewitt, 2015](#); [von der Embse et al., 2016](#)), we hypothesized that perceived pressure associated with teacher evaluation based on student performance would have a direct and negative effect on vitality and a direct and positive effect on exhaustion (H1). Second, based on SDT and previous research that has shown that controlling environments can undermine motivation (e.g., [Finnigan & Gross, 2007](#); [Yuan et al., 2012](#)), it was hypothesized that perceived pressure related to student performance evaluation would negatively predict autonomous motivation and positively predict controlled motivation and amotivation (H2). Third, in line with SDT and previous research (e.g., [Eyal & Roth, 2011](#); [Fernet et al., 2012](#)), we anticipated that autonomous motivation would positively predict vitality (an indicator of well-being) and negatively predict exhaustion (an indicator of ill-being), whereas controlled motivation and amotivation would negatively predict vitality and positively predict exhaustion (H3). Finally, we hypothesized that autonomous and controlled motivation, as well as amotivation, would significantly mediate the relations between pressure due to student performance-based evaluation and teacher vitality and exhaustion (H4; e.g., [Berliner, 2014](#); [Konstantopoulos, 2014](#)).

2. Method

2.1. Participants and educational context

A total of 360 Caucasian physical education teachers of secondary education (230 men and 130 women) from all regions of Spain participated in the study. The participants were between 23 and 61 years of age ($M = 40.51$; $SD = 9.01$), their work experience ranged between 1 and 39 years ($M = 14.57$ years, $SD = 9.62$ years), and they were employed in public ($n = 335$) and private ($n = 25$) schools. In the Spanish educational system, physical education is a mandatory subject and it has similar curricular structure, academic goals, and assessment systems to those of other subjects ([Pastor, Arribas, & Aguado, 2016](#)). Further, physical education teachers have the same staff development and promotion opportunities as other teachers. Alongside teaching a number of sports and games, the physical education curriculum requires the teaching of anatomy and physiology related to the study of general health and physical fitness.

The Spanish educational system has a curricular model focused on competency development. That is, all subjects, including physical education, should contribute to the development of key competencies (motor, linguistic, mathematics, digital, social, cultural, learning to learn, or entrepreneurial initiative). Students are evaluated through different methods, such as practical tests, teacher observations, written examinations and homework assignments. Through these methods, the teacher evaluates and grades the students using the criteria and standards established by the Ministry of Education. In addition, the physical education grades received by the students contribute to their global academic record ([Organic Law for the Improvement of the Educative Quality, 2013](#)).

2.2. Procedure

Schools and professional associations of physical education teachers were contacted and informed about the research objectives. These institutions approved and supported the project and facilitated contact with the participants via email. Following ethics approval from a Spanish university, consent from all participants was obtained. Participants were informed that the study was voluntary and their responses would be kept anonymous. The questionnaire was completed online and the measures were completed in the same order by all participants.

2.3. Measures

2.3.1. Perceived pressure due to student performance

A subscale of the Pressure at Work scale (Pelletier, Séguin-Lévesque, & Legault, 2002; Taylor et al., 2008), adapted to the Spanish context (Bartholomew, Ntoumanis, Cuevas, & Lonsdale, 2014), was used to assess the pressure perceived by teachers for their students to display strong academic performance. The Pressure for Evaluation based on Student Performance subscale consists of four items that measure the pressure perceived by the teacher when being evaluated according to the performance of his or her students (e.g., “My school will evaluate me poorly if my students don't get good grades”). The response range was 1 (*not at all true*) to 7 (*very true*). The scale has been used previously with physical education teachers in both Spanish (Bartholomew et al., 2014) and English (Taylor et al., 2008), with evidence of adequate reliability (Cronbach alphas of 0.79 and 0.75, respectively) and validity ($\chi^2(98) = 268.3, p < 0.001, CFI = 0.91, TLI = 0.90, RMSEA = 0.07$; Bartholomew et al., 2014).

2.3.2. Motivation

The Work Motivation Inventory (Blais, Brière, Lachance, Riddle, & Vallerand, 1993), adapted to the Spanish educational context (Cuevas, Sánchez-Oliva, Contreras, Moreno, & García-Calvo, 2014), was used to measure the different types of teacher motivation outlined within SDT (Deci & Ryan, 1985). The heading “Why do you teach?” was followed by six four-item subscales tapping each regulation: intrinsic motivation (e.g., “For the intense moments of pleasure teaching gives me”), integrated motivation (e.g., “Teaching is part of my life”), identified motivation (e.g., “I want to pursue my career in teaching”), introjected motivation (e.g., “I want to succeed at teaching, if not I would be very ashamed of myself”), external motivation (e.g., “For the income it provides me”), and amotivation (e.g., “I don't know, I have the impression that I don't have what it takes to teach”). The response scale used ranged from 1 (*not at all true*) to 7 (*very true*) scale. Following the SDT conceptualization (Deci & Ryan, 1985), the scores from the intrinsic, integrated and identified motivation subscales were averaged to form a single dimension termed “autonomous motivation”. In addition, the introjected and external motivation subscales were averaged to form a single variable termed “controlled motivation”. This combination of subscales has been previously used in a number of studies with strong psychometric evidence (e.g., autonomous motivation $\alpha = 0.85$, controlled motivation $\alpha = 0.76$; Vansteenkiste, Lens, De Witte, De Witte, & Deci, 2004). In addition, Cuevas et al. (2014) found adequate reliability ($\alpha > 0.76$ for the intrinsic, controlled and amotivation subscales) and validity ($\chi^2(234) = 780.91, p < 0.01, CFI = 0.93, TLI = 0.96, RMSEA = 0.08$) for the Spanish version of the scale.

2.3.3. Vitality

The Spanish adaptation (Balaguer, Castillo, Álvarez, & Duda, 2005) of the Subjective Vitality Scale (SVS; Ryan & Frederick, 1997) was used to measure the feeling of being full of energy and alive. The instrument consists of six items (e.g., “I feel alive and full of vitality”) that are assessed using a 1 (*not at all true*) to 7 (*very true*) scale. Previous studies have reported adequate factorial structure ($\chi^2(8) = 19.95, p < 0.01, CFI = 0.97, NFI = 0.95, RMSEA = 0.08$; Bostic, Rubio, & Hood, 2000) for the English version, and adequate reliability ($\alpha > 0.84$; Álvarez, Balaguer, Castillo, & Duda, 2012; Bostic et al., 2000) for the Spanish and English versions of the scale.

2.3.4. Exhaustion

The Spanish adaptation of the Exhaustion subscale (Gil-Monte, 2002) within the reduced version of the Maslach Burnout Inventory (MBI; Schaufeli, Leiter, Maslach, & Jackson, 1996) was used to assess mental exhaustion among the participants. It consists of five items designed to assess the mental fatigue and the decreased emotional resources of the participants (e.g., “Because of my job, I am exhausted”). The response range was 1 (*not at all true*) to 7 (*very true*). Gil-Monte (2002) and Bartholomew et al. (2014) reported evidence that supported the reliability ($\alpha = 0.84$ and 0.86 , respectively) and validity ($\chi^2(101) = 333.17, p < 0.01, CFI = 0.92, TLI = 0.91, RMSEA = 0.08$; Bartholomew et al., 2014) of this scale in Spanish populations.

2.4. Data analysis

Preliminary analyses were conducted using IBM-SPSS 20.0. In addition to Cronbach's alpha, composite reliability (CR) and average variance extracted (AVE) were calculated. CR indicates the degree of consistency of the observed variables with the measurement latent construct. AVE indicates the variance of the items captured by the latent construct compared to the variance captured by measurement error. Hair, Black, Babin, and Anderson (2010) considered acceptable values to be if CR is higher or equal to 0.07 and if AVE is higher or equal to 0.05. Means, standard deviations and bivariate correlations were also estimated.

Confirmatory factor analysis (CFA) and structural equation modeling were performed using AMOS 18.0. Factorial validity was tested with a CFA of the measurement model. For the structural equation modeling analysis, latent factors that correspond to pressure, amotivation, vitality, and exhaustion were estimated using the items from each scale as indicators. The latent factor for autonomous motivation was estimated based on the average values of intrinsic motivation and integrated and identified regulation. The latent factor for controlled motivation was estimated based on the averages of introjected and external regulation. Due to lack of normality in the data, the maximum-likelihood estimation method with bootstrapping was used. Bootstrapping provides robust standard errors estimates in the absence of normality (Byrne, 2001). The following indices were used to interpret model fit: the chi-square value, the comparative fit index (CFI), the Tucker-Lewis index (TLI), and the root mean square error of approximation (RMSEA). A model may be considered to be acceptable if CFI and TLI are close to or exceed 0.95 and if RMSEA is less than or equal to 0.08 (Hooper, Coughlan, & Mullen, 2008; Hu & Bentler, 1999; Tabachnick & Fidell, 2007). Additionally, with the RMSEA value, a

Table 1
Descriptive Statistics, Reliability Estimates and Pearson Correlations.

Variable	Range	M	SD	α	CR	AVE	1	2	3	4	5
1. Perceived pressure	1–7	2.78	0.92	0.72	0.70	0.52					
2. Autonomous motivation	1–7	5.31	0.96	0.87	0.82	0.62	0.03				
3. Controlled motivation	1–7	4.24	1.05	0.75	0.73	0.53	0.15**	0.48**			
4. Amotivation	1–7	2.38	1.04	0.67	0.65	0.50	0.33**	– 0.31**	0.05		
5. Vitality	1–7	5.55	1.22	0.94	0.93	0.69	– 0.25**	0.41**	0.08	– 0.44**	
6. Exhaustion	1–7	2.75	1.45	0.90	0.89	0.65	0.39**	– 0.30**	0.02	0.53**	– 0.61**

Note. α = Cronbach's alpha; CR = Reliability composite; AVE = Average variance extracted.

** $p < 0.01$.

confidence interval (i.e., 90%) is generated to indicate the level of the RMSEA precision. Quintana and Maxwell (1999) consider model fit to be adequate if the upper limit of this confidence interval is below 0.08 and if the range of the interval is smaller than 0.05.

Finally, direct and indirect effects (mediation analysis) were calculated using the bootstrapping technique proposed by Preacher and Hayes (2008) via IBM-SPSS 20.0. Bootstrapping generates a confidence interval (e.g., 95%) for indirect effects; if zero is included in the confidence interval, then the indirect effect is considered to be non-significant.

3. Results

3.1. Preliminary analyses

The CFA results supported the validity of the measurement model: ($\chi^2(150) = 259.63, p < 0.01, CFI = 0.97, TLI = 0.96, RMSEA(90\% CI) = 0.049(0.040–0.058)$). The means, standard deviations, Cronbach's alphas, composite reliabilities, and average mean extracted for each factor are presented in Table 1. On average, participants reported low mean levels of perceived pressure (2.78), amotivation (2.38), and exhaustion (2.75), whereas they reported high levels of autonomous motivation (5.31) and vitality (5.55). The reliability estimates were satisfactory (i.e. > 0.70) for all of the variables except amotivation, for which Cronbach's alpha (0.67) and composite reliability (0.65) were marginally acceptable (Hair et al., 2010). Table 1 also includes bivariate correlations, which were mostly in line with theoretical predictions in that perceived pressure from evaluations based on student performance was positively associated with controlled motivation (a weak association), amotivation (a moderate association), and exhaustion (a moderate association), and was negatively but relatively weakly associated with vitality. Moreover, amotivation was positively and relatively strongly associated with exhaustion and negatively and moderately associated with vitality whilst opposite relations were observed between autonomous motivation and exhaustion (i.e., a moderate negative association) and vitality (i.e., a moderate positive association).

3.2. Structural equation model

The hypothesized model (Fig. 1) assumed that pressure due to evaluation would negatively predict autonomous motivation, whereas it would positively predict controlled motivation and amotivation. In addition, autonomous motivation would positively predict vitality and negatively predict exhaustion, whereas controlled motivation and amotivation would positively predict exhaustion and negatively predict vitality. The model demonstrated acceptable fit indices ($\chi^2(168) = 364.19, p < 0.01, CFI = 0.95, TLI = 0.94, RMSEA = 0.06(90\% CI = 0.05–0.06)$). All of the hypothesized relations were significant ($p < 0.01$) except that between perceived pressure and controlled motivation, which was not significant (providing partial support for H2 & H3).

3.3. Direct and indirect effects

Table 2 presents the direct effects of pressure due to evaluation on vitality and exhaustion, and the indirect effects of pressure on vitality and exhaustion through the motivation variables. Perceived pressure directly and positively predicted exhaustion and negatively predicted vitality (providing support of H1). The total indirect effects were significant. When examining the specific indirect effects, it was observed that these effects were not significant for autonomous and controlled motivation. In contrast, amotivation mediated the negative indirect effect from pressure on vitality, and the positive indirect effect of pressure on exhaustion (providing partial support for H4).

4. Discussion

The primary purpose of the current study was to test a model, based on SDT, that examined whether the pressure experienced by physical education teachers when evaluated based on the performance of their students was related to different types of motivation for teaching and, in turn, to teacher psychological well-being (vitality) and ill-being (exhaustion). The mediating role of autonomous motivation, controlled motivation, and amotivation in the relationship between perceived pressure and both vitality and exhaustion was also tested. There is a dearth of studies linking teacher evaluations based on student performance with teacher motivation, well-

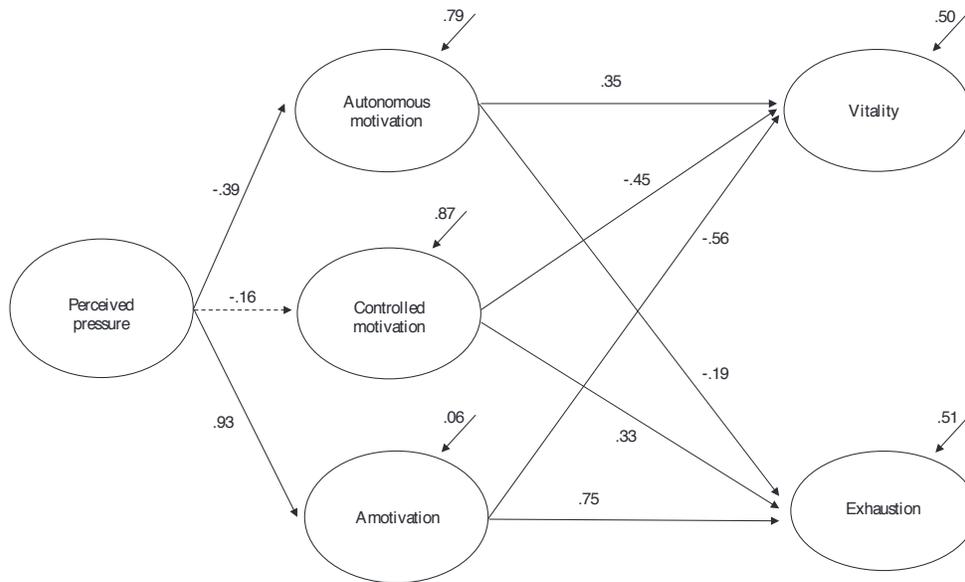


Fig. 1. Final model for the prediction of teacher well-being and teacher ill-being.

Note: Significant paths are marked by solid lines. The small arrows over the dependent variables represent residual variance.

Table 2
Standardized direct and indirect effects.

Independent variable	Criterion variable	Specific indirect effects				
		Total direct effect (95% CI)	Total indirect effect (95% CI)	Autonomous motivation (95% CI)	Controlled motivation (95% CI)	Amotivation (95% CI)
Perceived pressure	Vitality	- 0.22* (0.34 to 0.97)	- 0.10* (0.20 to 0.02)	0.02 (- 0.04 to 0.07)	- 0.01 (- 0.03 to 0.01)	- 0.11* (- 0.18 to - 0.06)
Perceived pressure	Exhaustion	0.33* (0.19 to 0.47)	0.19* (0.10 to 0.31)	- 0.01 (- 0.06 to 0.03)	0.01 (- 0.02 to 0.04)	0.20* (0.13 to 0.30)

Note. CI = 95% Confidence Intervals, * = CI does not include zero.

being, and ill-being. Hence, this study offers important empirical evidence regarding the effects of this type of teacher evaluation.

The first hypothesis that perceived pressure related to teacher evaluation based on student performance would negatively predict vitality and positively predict exhaustion was fully supported. The direct effects of perceived pressure on vitality and exhaustion were significant and in the expected direction. These results were consistent and comparable in size with previous findings that associated the pressure due to this type of teacher evaluation with stress (von der Embse et al., 2016), anxiety (Goldhaber & Hannaway, 2004; Hewitt, 2015), and burnout (Dworkin & Tobe, 2014) among teachers. Given that teacher well-being has been linked with better teaching performance (Klusmann, Kunter, Trautwein, Lüdtke, & Baumert, 2008; Roth et al., 2007), whereas teacher ill-being has been linked to negative teacher-student relationships (von der Embse et al., 2016) and poorer student academic performance (Blandford, 2000), these findings suggest that evaluations based on student performance should be implemented very cautiously.

The second hypothesis, which proposed that perceived pressure related to student performance evaluation would negatively predict autonomous motivation and positively predict controlled motivation and amotivation, was partially supported. Perceived pressure negatively predicted teacher autonomous motivation. That is, the greater pressure that teachers felt from performance-based evaluations, the less likely they were to report that they taught for reasons of interest and personal value. These results align with those of previous studies that have reported a negative association between pressure due to teaching assessment and internal (i.e., more self-determined) forms of motivation (Finnigan & Gross, 2007; Yuan et al., 2012). In addition, we found that pressure due to teacher evaluation based on student performance also positively predicted amotivation among teachers. In other words, the greater pressure that teachers felt from performance-based evaluations, the more likely they were to report a complete absence of motivation to teach (i.e., their decision to teach was neither intrinsically nor extrinsically motivated). Taken together, these findings suggest that perceived pressure due to teacher evaluations based on student performance could undermine autonomous motivation and promote teacher amotivation. Motivational deficits could promote teacher ill-being (Eyal & Roth, 2011) which, in turn, could adversely affect the quality of the teacher's professional work (Klusmann et al., 2008).

However, it must also be noted that, contrary to what was hypothesized, pressures associated with this type of student-based teacher evaluation did not significantly predict controlled motivation in teachers. It is likely that the threats associated with the punitive aspects of this method of evaluation were perceived to be far greater than any rewards that would be available from improved student performance (Marina et al., 2015). For example, the consequences of poor teacher evaluations in Spain include

difficulties in career progression, a reduction in responsibilities and status, and even job loss (Ministry of Education, Culture and Sport, 2014). However, the scale assessing external regulation (one of the two components of controlled motivation) used in the present study mainly captures positive rewards (e.g., economic gains and job security), which could explain the absence of relation between perceived pressure and controlled motivation. Hence, future studies in this area should ensure that the assessment of external regulation captures both punishments and rewards. Introjected regulation refers to internal pressures; hence, it is less likely that this component of controlled motivation would be predicted by external pressures associated with student-based teacher evaluation.

The third hypothesis, which concerned the relations between motivation, well-being, and ill-being, was fully supported. In line with SDT, autonomous motivation positively predicted vitality and negatively predicted exhaustion. That is, participants who reported that they taught for pleasure or personal growth were more likely to report feeling full of life and less likely to report feeling exhausted. These findings further underscore the importance of motivation based on autonomous factors, such as interest or value, for nurturing teacher well-being and psychological health (Eyal & Roth, 2011; Fernet et al., 2012; Roth et al., 2007). This is important because autonomously motivated teachers facilitate supportive teaching environments which, in turn, promote students' autonomous motivation for learning (Roth et al., 2007). On the other hand, controlled motivation and amotivation were negatively associated with vitality and positively associated with exhaustion. These results are also in line with other findings (Firestone, 2014; Yuan et al., 2012), which have indicated the limited effectiveness of initiatives aimed at improving teacher motivation based on extrinsic factors (e.g., higher payment for teachers who achieve better student performance) without considering the internal motivational resources of teachers.

The fourth hypothesis of our study posited that the different types of motivation would play a mediating role between perceived pressure due to student performance and teacher well-being and ill-being. This hypothesis was only partially supported. Specifically, the total indirect effects of perceived pressure on vitality and exhaustion were significant. However, an analysis of the specific indirect effects indicated that only amotivation (which, in contrast to autonomous and controlled motivation, represents the complete lack of motivation) played a mediating role in the relations between pressure and vitality, and between pressure and exhaustion. In other words, perceived pressure resulted in higher levels of exhaustion and lower levels of vitality via increased amotivation. These results indicate that mediation only occurs in the absence of motivation (amotivation). That is to say, perceived pressure predicts low vitality and high exhaustion because it makes teachers experience a sense of helplessness. It is possible that controlled motivation was not a significant mediator because the questionnaire assessing the external regulation component of controlled motivation focused primarily on the rewards aspect of control (e.g., financial gains). It would have been beneficial to also assess the punishment aspect of external regulation (e.g., penalties or delays in career progression), which is more likely to be associated with this type of teacher evaluation. In addition, from a statistical perspective, controlled and autonomous motivation may not have been significant mediators because the direct effects from perceived pressure to amotivation and, in turn, from amotivation to the two dependent variables were very strong, leaving little unique variance for autonomous and controlled motivation to account for. Such findings are particularly useful because they contribute to the understanding of potential mechanisms through which pressure due to student performance-contingent teacher evaluation can affect teacher well-being and ill-being.

4.1. Limitations and directions for future research

This study has a number of limitations that should be considered in future research. First perceived pressure from student performance-contingent evaluation was fairly low in the current sample ($M = 2.78$, $SD = 0.92$), perhaps due to the subject that these teachers were teaching. The observed results may be different in a sample in which these pressures were felt more intensely by teachers. Second, the sample of teachers was confined to a single country, a single subject, and a single academic level. Therefore, the generalizability of the current findings to other countries, school subjects, and academic levels is unknown and should be explored. Third, the study was cross-sectional in nature, which prevented us from testing causal relations between variables. Consequently, new experimental studies could complement the results of our study by comparing the effects of different types of teacher evaluation on the well-being/ill-being of teachers. Fourth, the current study focused narrowly on how altered teacher motivation due to perceived evaluation pressures affects teachers' reported levels of vitality and exhaustion; however, there are many other dependent variables that could be explored, including objective records of teachers' health, teachers' turnover decisions, teachers' interpersonal behaviors and student motivation or student engagement. Finally, the results concerning amotivation should be viewed with caution, as the reliability index of this measure, the only significant mediator in this study, was marginally under the recommended cut-off value of 0.70.

Previous research (Berliner, 2014; Firestone, 2014; Weiner, 2004; Yuan et al., 2012) has suggested that decreases in teacher motivation and well-being could be related to the fact that this type of teacher evaluation focuses on metrics that are outside of the direct control of the teacher (e.g., the performance of their students). As such, an important direction for future research would be to explore whether teacher evaluation based on controllable (e.g., class preparation or teaching skills) or non-controllable aspects (e.g., student performance) differently affect motivation and well-being of teachers. It may be that more controllable types of evaluation can counteract the negative effects of less controllable ones by giving teachers opportunities to demonstrate their competencies. In addition, some types of evaluation only consider student performance in particular subjects (e.g., Math and reading). It would, therefore, be interesting to compare how this assessment affects teachers' psychological health depending on whether their subjects are included in the evaluation or not.

4.2. Implications for practice

The results of the current study have demonstrated how teacher evaluation based on student performance can negatively affect teacher well-being, which has several potential implications for applied practice. One potential alternative to conducting teacher evaluation based on student performance would be to adopt a more holistic and inclusive definition of what is meant by good teaching. Good and Lavigne (2015), for example, emphasized that good teaching involves much more than increasing students' grades; for example, it means supporting students to become better problem solvers (promoting creativity and analytical skills) or stimulating students' civility and social responsibility (promoting respect and empathy). In other words, besides the traditional grades, teacher evaluations could also incorporate information about the degree of development of cognitive and social skills of students. As the results of the present study indicate that teacher evaluation based on student performance can negatively affect motivation and teacher well-being, school should consider the leadership styles of principals and administration managers with the emphasis being on helping them minimize coercive strategies, such as rewards and comparisons with others, and instead promoting autonomous teacher motivation for work (Eyal & Roth, 2011; Fernet et al., 2012).

References

- Álvarez, M. S., Balaguer, I., Castillo, I., & Duda, J. L. (2012). The coach-created motivational climate, young athletes' well-being, and intentions to continue participation. *Journal of Clinical Sport Psychology*, 6, 166–179. <http://dx.doi.org/10.1123/jcsp.6.2.166>.
- Baker, E. L., Barton, P. E., Darling-Hammond, L., Haertel, E., Ladd, H. F., Linn, R. L., ... Shepard, L. A. (2010). Problems with the use of student test scores to evaluate teachers. *EPI briefing paper*. Vol. 278 Economic Policy Institute.
- Balaguer, I., Castillo, I., Álvarez, M., & Duda, J. L. (2005). Importance of social context in the prediction of self-determination and well-being in athletes of different level. *Symposium on motivation in sport and physical activity*. CD-Rom of 9th European Congress of Psychology. Granada, Spain.
- Bartholomew, K. J., Ntoumanis, N., Cuevas, R., & Lonsdale, C. (2014). Job pressure and ill-health in physical education teachers: The mediating role of psychological need thwarting. *Teaching and Teacher Education*, 37, 101–107. <http://dx.doi.org/10.1016/j.tate.2013.10.006>.
- Berliner, D. C. (2014). Exogenous variables and value-added assessments: A fatal flaw. *Teachers College Record*, 116, 1–31.
- Blais, M. R., Brière, N. M., Lachance, L., Riddle, A. S., & Vallerand, R. J. (1993). L'inventaire des motivations au travail de Blais [Blais' inventory of motivations at work]. *Revue québécoise de psychologie*, 14(3), 185–215.
- Blandford, S. (2000). *Managing professional development in schools*. London, UK: Routledge.
- Bostic, T. J., Rubio, D. M., & Hood, M. (2000). A validation of the subjective vitality scale using structural equation modeling. *Social Indicators Research*, 52, 313–324. <http://dx.doi.org/10.1023/A:1007136110218>.
- Byrne, B. M. (2001). *Structural equation modeling with AMOS: Basic concepts, applications, and programming*. Mahwah, NJ: L Erlbaum Associates.
- Cuevas, R., Sanchez-Oliva, D., Bartholomew, K. J., Ntoumanis, N., & Garcia-Calvo, T. (2015). Adaptation and validation of the Psychological Need Thwarting Scale in Spanish physical education teachers. *Spanish Journal of Psychology*, 18, 1–9. <http://dx.doi.org/10.1017/sjp.2015.56>.
- Cuevas, R., Sánchez-Oliva, D., Contreras, O., Moreno, H., & García-Calvo, T. (2014). Adaptación del inventario de motivación en el trabajo al contexto educativo español [Adaptation of the motivational inventory at work to the Spanish educational context]. In A. Romero, T. Ramiro-Sánchez, & M. P. Bermúdez (Eds.). *Libro de resúmenes de II Congreso Internacional de Ciencias de la Educación y del Desarrollo*. Granada: Asociación de Psicología Conductual.
- Deci, E. L., & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior*. New York: Academic Press.
- Dworkin, A. G., & Tobe, P. F. (2014). The effects of standards based school accountability on teacher burnout and trust relationships: A longitudinal analysis. In D. V. Maele, P. B. Forsyth, & M. V. Houtte (Eds.). *Trust and school life: The role of trust for learning, teaching, leading, and bridging* (pp. 121–143). Netherlands: Springer.
- von der Embse, N. P., Pendergast, L. L., Segool, N., Saeki, E., & Ryan, S. (2016). The influence of test-based accountability policies on school climate and teacher stress across four states. *Teaching and Teacher Education*, 59, 492–502. <http://dx.doi.org/10.1016/j.tate.2016.07.013>.
- Eyal, O., & Roth, G. (2011). Principals' leadership and teachers' motivation: Self-determination theory analysis. *Journal of Educational Administration*, 49, 256–275. <http://dx.doi.org/10.1108/09578231111129055>.
- Feng, L., Figlio, D. N., & Sass, T. (2010). *School accountability and teacher mobility*. Cambridge, USA: National Bureau of Economic Research.
- Fernet, C., Guay, F., Senécal, C., & Austin, S. (2012). Predicting intraindividual changes in teacher burnout: The role of perceived school environment and motivational factors. *Teaching and Teacher Education*, 28, 514–525. <http://dx.doi.org/10.1016/j.tate.2011.11.013>.
- Finnigan, K. S., & Gross, B. (2007). Do accountability policy sanctions influence teacher motivation? Lessons from Chicago's low-performing schools. *American Educational Research Journal*, 44, 594–630. <http://dx.doi.org/10.3102/0002831207306767>.
- Firestone, W. A. (2014). Teacher evaluation policy and conflicting theories of motivation. *Educational Researcher*, 43, 100–107. <http://dx.doi.org/10.3102/0013189X14521864>.
- Gil-Monte, P. R. (2002). Validez factorial de la adaptación al español del Maslach Burnout Inventory-General Survey [The factorial validity of the Maslach Burnout Inventory-General Survey]. *Salud Pública de México*, 44, 33–40.
- Goldhaber, D., & Hannaway, J. (2004). Accountability with a kicker: Observations on the Florida A+ accountability plan. *Phi Delta Kappan*, 85, 598–605. <http://dx.doi.org/10.1177/003172170408500807>.
- Good, T. L., & Lavigne, A. L. (2015). Issues of teacher performance stability are not new: Limitations and possibilities. *Education Policy Analysis Archives*, 23, 1–16. <http://dx.doi.org/10.14507/epaa.v23.1916>.
- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate data analysis*. New Jersey, USA: Maxwell Macmillan: International Editions.
- Harvey, L. (2004). The power of accreditation: Views of academics. *Journal of Higher Education Policy and Management*, 26, 207–223. <http://dx.doi.org/10.1080/1360080042000218267>.
- Hewitt, K. K. (2015). Educator evaluation policy that incorporates EVAAS value-added measures: Undermined intentions and exacerbated inequities. *Education Policy Analysis Archives*, 23, 1–49. <http://dx.doi.org/10.14507/epaa.v23.1968>.
- Hooper, D., Coughlan, J., & Mullen, M. (2008). Structural equation modelling: Guidelines for determining model fit. *Electronic Journal of Business Research Methods*, 6, 53–60.
- Hu, L., & Bentler, P. M. (1999). Cut-off criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6, 1–55. <http://dx.doi.org/10.1080/10705519909540118>.
- Isore, M. (2009). Teacher evaluation: current practices in OECD countries and a literature review. *OECD education working papers*. Vol. 23 OECD Publishing (NJ1).
- Jiang, J. Y., Sporte, S. E., & Luppescu, S. (2015). Teacher perspectives on evaluation reform Chicago's REACH students. *Educational Researcher*, 44, 105–116. <http://dx.doi.org/10.3102/0013189X15575517>.
- Klusmann, U., Kunter, M., Trautwein, U., Lüdtke, O., & Baumert, J. (2008). Teachers' occupational well-being and quality of instruction: The important role of self-regulatory patterns. *Journal of Educational Psychology*, 100, 702–715. <http://dx.doi.org/10.1037/0022-0663.100.3.702>.
- Konstantopoulos, S. (2014). Teacher effects, value-added models, and accountability. *Teachers College Record*, 116, 1–21.
- Marina, J. A., Pellicer, C., & Manso, J. (2015). *Libro blanco de la profesión docente y su entorno escolar [White paper of the teaching profession and its school environment]*. Madrid, Spain: Ministerio de Educación, Cultura y Deporte.
- McCaffrey, D. F., Lockwood, J. R., Koretz, D. M., & Hamilton, L. S. (2003). *Evaluating value-added models for teacher accountability*. Monograph. PO Box 2138, Santa

- Monica, CA 90407-2138: RAND Corporation.
- McCaffrey, D. F., Sass, T. R., Lockwood, J. R., & Kata, M. (2009). The intertemporal variability of teacher effect estimates. *Education Finance and Policy*, 4, 572–606. <http://dx.doi.org/10.1162/edfp.2009.4.4.572>.
- Ministry of Education, Culture, and Sport (2014). *TALIS 2013 International study of teaching and learning. Spanish Report*. Madrid: Subdirección General de Documentación y Publicaciones.
- Organic Law for the Improvement of the Educative Quality 8/2013 (2013). 9th December *Ley Orgánica para la Mejora de la Calidad Educativa* Vol. 295. Spanish State Official Bulletin, [Boletín oficial del Estado de España].
- Organisation for Economic Co-operation and Development (2014). *New insights from TALIS 2013. Teaching and learning in primary and upper secondary education* Paris, France: OECD Publishing. <http://dx.doi.org/10.1787/9789264226319-en>.
- Pas, E. T., Bradshaw, C. P., & Hershfeldt, P. A. (2012). Teacher- and school-level predictors of teacher efficacy and burnout: Identifying potential areas for support. *Journal of School Psychology*, 50, 129–145. <http://dx.doi.org/10.1016/j.jsp.2011.07.003>.
- Pastor, V. M. L. D. P., Arribas, J. C. M., & Aguado, R. M. (2016). Los retos de la Educación Física en el Siglo XXI [The challenges of physical education in the 21st century]. *Retos*, 29, 182–187.
- 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. <http://dx.doi.org/10.1037/0022-0663.94.1.186>.
- Preacher, K. J., & Hayes, A. F. (2008). Contemporary approaches to assessing mediation in communication research. In A. F. Hayes, M. D. Slater, & L. B. Snyder (Eds.). *The Sage sourcebook of advanced data analysis methods for communication research* (pp. 13–54). Thousand Oaks, CA: Sage Publications. <http://dx.doi.org/10.4135/9781452272054.n2>.
- Quintana, S. M., & Maxwell, S. E. (1999). Implications of recent developments in structural equation modeling for counseling psychology. *The Counseling Psychologist*, 27, 485–527. <http://dx.doi.org/10.1177/0011000099274002>.
- Rockoff, J. E. (2004). The impact of individual teachers on student achievement: Evidence from panel data. *The American Economic Review*, 94, 247–252.
- Roth, G., Assor, A., Kanat-Maymon, Y., & Kaplan, H. (2007). Autonomous motivation for teaching: How self-determined teaching may lead to self-determined learning. *Journal of Educational Psychology*, 99, 761–774. <http://dx.doi.org/10.1037/0022-0663.99.4.761>.
- Rothstein, J. (2010). Teacher quality in educational production: Tracking, decay, and student achievement. *Quarterly Journal of Economics*, 125, 175–214. <http://dx.doi.org/10.1162/qjec.2010.125.1.175>.
- Ryan, R. M., & Frederick, C. (1997). On energy, personality, and health: Subjective vitality as a dynamic reflection of well-being. *Journal of Personality*, 65, 529–565. <http://dx.doi.org/10.1111/j.1467-6494.1997.tb00326.x>.
- Schaufeli, W. B., Leiter, M. P., Maslach, C., & Jackson, S. E. (1996). The MBI-general survey. In C. Maslach, S. E. Jackson, & M. P. Leiter (Eds.). *Maslach Burnout inventory. Manual* (pp. 19–26). Palo Alto, CA: Consulting Psychologists Press.
- Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). New York: Allyn and Bacon.
- Taylor, E. S., & Tyler, J. H. (2012). The effect of evaluation on teacher performance. *American Economic Review*, 102, 3628–3651. <http://dx.doi.org/10.1257/aer.102.7.3628>.
- Taylor, I., & Ntoumanis, N. (2007). Teacher motivational strategies and student self-determination in physical education. *Journal of Educational Psychology*, 99, 747–760. <http://dx.doi.org/10.1037/0022-0663.99.4.747>.
- Taylor, I., Ntoumanis, N., & Standage, M. (2008). A self-determination theory approach to understanding antecedents of teachers' motivational strategies in physical education. *Journal of Sport and Exercise Psychology*, 30, 75–94. <http://dx.doi.org/10.1123/jsep.30.1.75>.
- Vansteenkiste, M., Lens, W., De Witte, S., De Witte, H., & Deci, E. L. (2004). The 'why' and 'why not' of job search behaviour: their relation to searching, unemployment experience, and well-being. *European Journal of Social Psychology*, 34, 345–363. <http://dx.doi.org/10.1002/ejsp.202>.
- Weiner, B. (2004). Attribution theory revisited: Transforming cultural plurality into theoretical unity. In D. M. McInerney, & S. Van Etten (Eds.). *Big theories revisited* (pp. 13–29). Greenwich, CT: Information Age Publishing.
- Yuan, K., Le, V. N., McCaffrey, D. F., Marsh, J. A., Hamilton, L. S., Stecher, B. M., & Springer, M. G. (2012). Incentive pay programs do not affect teacher motivation or reported practices results from three randomized studies. *Educational Evaluation and Policy Analysis*, 35, 3–22. <http://dx.doi.org/10.3102/0162373712462625>.