



University of Groningen

Validating a model of effective teaching behaviour and student engagement

Inda-Caro, Mercedes; Maulana, Ridwan; Fernández-García, Carmen María; Peña-Calvo, José Vicente; Rodríguez-Menéndez, M. del Carmen; Helms-Lorenz, Michelle

Published in:
Learning Environments Research

DOI:
[10.1007/s10984-018-9275-z](https://doi.org/10.1007/s10984-018-9275-z)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Final author's version (accepted by publisher, after peer review)

Publication date:
2019

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Inda-Caro, M., Maulana, R., Fernández-García, C. M., Peña-Calvo, J. V., Rodríguez-Menéndez, M. D. C., & Helms-Lorenz, M. (2019). Validating a model of effective teaching behaviour and student engagement: perspectives from Spanish students. *Learning Environments Research*, 22(2), 229–251.
<https://doi.org/10.1007/s10984-018-9275-z>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

VALIDATING A MODEL OF EFFECTIVE TEACHING BEHAVIOUR AND STUDENT ENGAGEMENT: PERSPECTIVES FROM THE SPANISH STUDENTS

1. Introduction

Interest in research into productive learning environment indicators has increased during the last century. Teaching behaviour, particularly, is recognized as a highly important indicator of learning environments. The importance of teaching behaviour for student outcomes has been highlighted in studies on teacher effectiveness indicating that classroom factors are more important than school factors, and teaching behaviour is a classroom factor that matters most (Muijs et al., 2014; Townsend, 2007). Consequently, improvement in teaching behaviour has been called upon to be included in teachers' initial training and teachers' professional development agenda. A number of studies have documented the productive functioning of teaching behaviour **domains** which contribute to productive learning environments (i.e. classroom environment, learning climate, class control, instructional support) leading to the improvement of students' affective and cognitive outcomes (Antoniou, Kyriakides, & Creemers, 2011; Centra & Potter, 1980; Guldemon & Bosker, 2009; Hattie, 2003; Konstantopoulos & Sun, 2014; Kyriakides & Creemers, 2009; Muijs, Campbell, Kyriakides, & Robinson, 2005; Opdenakker & Van Damme, 2001, 2006; Opdenakker, VanDamme, De Fraine, Van Landeghem, & Onghena, 2002; Teodorovic, 2011; Van den Broeck, Opdenakker, & Van Damme, 2005).

Thus far, the actions developed by teachers during their teaching **training** have been one of the central **domains** in this area of research, assuming that better teachers can only be identified after some evidence on their actual job performance (Staiger & Rockoff, 2010). Teaching behaviour is generally viewed to be multidimensional in nature (Burdsal & Bardo, 1986; Guskey & Passaro, 1994; Muijs et al., 2005). Nevertheless, there is no consensus concerning the most suitable term to refer to the best teaching **behaviours**, the number and nature of the **domains** or the most appropriate way to assess them (Burdsal & Bardo, 1986; Maulana, Helms-Lorenz, Van de Grift, 2017).

Some countries have developed systems to review the extent to which a teacher has contributed to student achievement gains in a school year but, according to Van der Lans, Van de Grift and Van Veen (2015) this value - added approach should be complemented with other evaluation methods. A student questionnaire, which is viewed as the most cost-effective method of classroom environment measure, can be used to capture a representative image of day to day teachers' behaviours (De Jong & Westerhof, 2001; Hoyt & Pallet, 1999; Opdenakker & Minnaert, 2011). In the Spanish context, particularly, there is a need for a cost-effective, highly reliable and valid measure of teaching behaviour as a means to support the teacher professional development agenda **continuously**.

The Spanish-speaking teaching contexts would benefit from instruments to measure effective teaching behaviour for several reasons. First, the study of teaching behaviour as a determinant of learning environments has been prolific. However, insights from the Spanish-speaking contexts are limited, while globally the population of Spanish-speaking people is remarkable. The validation of a Spanish instrument would contribute to provide more insights about teaching behaviour from various Spanish-speaking contexts. Second, results from international testing studies, such as Trends in International Mathematics and Science Study (TIMSS) and Program for International Student Assessment (PISA), regarding differences in student achievement around the globe have motivated researchers to search for explanations from the classroom level in terms of teaching behaviour. This is informed by the teaching effectiveness literature that, besides student-level factors, teacher factors explain a considerable amount of variance in student achievement (e.g., Aaronson, Barrow, & Sander, 2007; Bosker & Witziers, 1996; Houtveen, Van de Grift, & Brokamp, 2014). Subsequently, many contemporary researchers are interested in comparing teaching behaviour internationally (e.g., Maulana et al., 2017; Van de Grift et al., 2017). Third, the present research is embedded within a

larger international study aiming among other objectives, to compare teaching behaviour across countries and to study the possibility of comparable international teaching behaviour profiles. Validating Spanish instruments to measure effective teaching behavior is the first step towards international comparison studies in teaching behaviour, that aim to contribute to advancing the knowledge base of productive learning environments from the teaching behaviour lens.

2. Teaching behaviour

Theories of teacher development have been largely studied resulting in different models which focus on several variables to understand and explain teachers' behaviour -for an exhaustive meta-analysis about some of these models and their research approach, see Scheerens (2016) and Seidel and Shavelson (2007)-. Since Newmark's (1929) study where he asked students to identify their best and poorly performing teachers, a significant evolution is visible in this field. Although the context of secondary education has changed considerably since the 1920s, some characteristics stated in Newmark's study remain valid nowadays (e.g. getting ideas across to students, cooperating with students, daily preparation of the lessons, showing interest in students, broad grasp of subject matter) and refer to learning climate factors and teacher – student interactions.

Years later, Fuller (1969) developed a study to investigate teachers' needs by using counseling seminars and written concerns statements. She described progressive changes in teacher concerns while they improved their professional experience, indicating a shift from concerns focused on the self (e.g. the limits of the acceptance in the institution) to others more directed to the task (e.g. cope with students' evaluation) then culminating in concerns about students and their teaching impact on them (e.g. ability to understand students' capacities).

Hattie (2003) identified five major domains which in terms of Fuller's classification refer to task and student impact concerns: identifying essential representations of their subject, guiding learning through classroom interactions, monitoring learning and providing feedback, attending affective attributes and influencing student outcomes.

With a broader perspective, Pianta and Hamre (2009) in their work in the Classroom Assessment Scoring System (CLASS) developed a standardized model of global classroom quality which assessed three basic domains of teaching: emotional supports, classroom organization and instructional supports. In this model, emotional supports refer to positive classroom climate, teacher sensitivity and regard for student perspectives. Classroom organization includes effective behaviour management, productivity and instructional learning formats. Finally, the instructional support domain considers the concept of development, quality of feedback and language modeling.

These approaches which connect different behaviours in domains, reinforce the idea that the improvement of teacher behaviour cannot be focused on the acquisition of isolated competencies but on helping teachers develop types of teacher behaviour that are more effective than others. They also stimulate reflection across the whole process of teaching in order to make teachers excellent practitioners (Kyriakides, Creemers, & Antoniou, 2009; Antoniou et al., 2011). With this idea in mind, Creemers introduces a dynamic model with eight factors referred to instructional behaviours of effective teaching (Kyriakides, Christoforou, & Charalambous, 2013; Kyriakides & Creemers, 2009; Kyriakides et al., 2009). All the factors included in this dynamic model (orientation, structuring, questioning, teaching modelling, application, management of time, teacher role in making classroom environment and classroom assessment) were measured in different domains which consider not only quantitative features but also more qualitative ones. Antoniou et al., (2011) assume that these factors and their domains may be interrelated, so they group the eight factors into five stages: basic elements of direct teaching, putting aspects of quality in direct teaching and touching on active teaching, acquiring quality in active/direct teaching, differentiation for teaching and finally, achieving quality and differentiation in teaching using different approaches. They also maintain that teacher improvement and stage growth do not unilaterally unfold but require a stimulating and supportive environment.

Taking into consideration the state of the art on teaching effectiveness and classroom environments research, as well as other models and empirical findings of research in teaching, Van de Grift (2007) introduced a model of observable teaching behaviour¹. This model is evidence – based and allows the study of the different stages of teaching skills throughout teachers’ professional career. According to this model, observable teaching behaviour can be distinguished into six teaching domains including *safe learning climate, efficient classroom management, clarity of instruction activating teaching, teaching – learning strategies, and differentiation* (Maulana et al., 2017; Van de Grift et al., 2014). These six domains conceptually overlap with other models of teaching behaviour reviewed above (Maulana, Helms-Lorenz, & Van de Grift, 2015a; Van de Grift et al., 2014), but measure distinct aspects of teaching sufficiently, and confirm a higher order latent construct called effective teaching behaviour.

In the present study, we used the model of teaching behaviour based on Van de Grift et al. (2014). This model provides conceptual clarity regarding the six effective teaching behaviour domains, which can be used as guidance for teacher professional development. Additionally, the teaching behaviour model has been proven to be valid (Maulana et al., 2017; Van de Grift et al., 2017). In the following section, the six domains are discussed more elaborately.

2.1 Safe learning climate

A safe learning climate requires the mutual respect not only between students and teachers but also among students to encourage students’ self – confidence and to facilitate good relationships in the classroom (Van de Grift et al., 2014; Maulana et al., 2015a, 2015b). Danielson’s (2013) evaluation instrument also refers to this domain as the creation of an environment of respect and rapport. According to Van de Grift (2007) these requisites make it possible to build an orderly and safe atmosphere in which students are stimulated to learn. Although not all studies carried out have a precise definition of educational climate, about 20% - 40% of the differences in students’ achievement could be explained by school climate factors including instruction or monitoring of students’ achievement (Van de Grift, 2007).

2.2. Efficient classroom management

Efficient classroom management presumes that the teacher is able to organize the learning time with behaviours such as avoiding the waste of time, punctuality in the beginning and ending of the lesson, providing well-structured classes, maximizing instructional time and avoiding students’ waiting for teachers’ attention (Danielson, 2013; Van de Grift, 2007, 2014; Van de Grift, Helms-Lorenz, & Maulana, 2014). Research indicates that more academically effective teachers have generally better organized classrooms and fewer behavioural problems with students (Van de Grift, 2007) and deal with students’ misbehavior more efficiently (Maulana, Helms-Lorenz, & Van de Grift, 2015b). Other important aspects are presenting information in an orderly manner and managing lesson and topic transitions accurately (Van de Grift, 2007; Maulana, Helms-Lorenz & Van de Grift, 2015a, 2017).

2.3. Clarity of instruction

Clarity of instruction includes a clear structure of the lesson, clarifying lesson objectives in order to let students know what they are expected to do during the lesson (Van de Grift, 2014; Maulana et al., 2015a), taking into account previous knowledge, giving clear examples, supervising the acquisition of objectives, the equilibrium of activities (dividing individual and group work clearly and in a balanced way) and offering immediate feedback to keep students on task, among others (Van de Grift et al., 2014; Maulana et al., 2015a, 2015b).

¹ Observable teaching behaviour refers to behaviours that are observable in the classroom such as a teacher giving instruction for students in the classroom. This type of behaviour can be observed by external individuals such as observers or students. This is distinguishable from other teaching behaviours that are not observable in the classroom such as a teacher planning a lesson (Maulana & Schuurman, 2018).

2.4. Activating teaching

Activating teaching entails connecting students' prior knowledge and the use of advance organizers (Van de Grift et al., 2014) so that contents make sense to students and let them be aware of the relevance of the lessons (Van de Grift, 2007; Maulana et al., 2015b). Recent studies have also shown that an activating learning environment is related to the quality of teacher – students and peer interactions (Maulana et al., 2015a). **When these relations improve, students' learning performances tend to improve as well (Anderson, Christenson, Sinclair, & Lehr, 2004; Furrer & Skinner, 2003).**

2.5. Teaching learnings strategies

Teaching learning strategies cover the use of scaffolds or other metacognitive strategies, which help students bridge the gap between the new concepts and the already known ones and to perform higher level procedures (Maulana et al., 2015a; Van de Grift, 2014, Van de Grift et al., 2014). They usually imply breaking problems down into more simple tasks that students have a real chance of solving (Van de Grift, 2007, 2014). **This domain has a clear conceptual relation with the factors developed by other scholars such as modelling in the dynamic model of educational effectiveness of Creemers & Kyriakides (2008) or Danielson (2013). Indeed they refer to the use of problem solving strategies, to promote the idea of modelling and that students may serve as resources for one another.**

2.6. Differentiation

Differentiation requires adapting teaching to student individual differences, demonstrating knowledge of students and addressing students' levels, learning preferences and learning profiles (Danielson, 2013; Maulana et al., 2015a). To prescribe and adapt the instructional methods for all students who may be identified at risk, students need to be rigorously diagnosed (Van de Grift, 2007). Several indicators reflect differentiated teaching behaviors: devoting extra time and additional instructions, pre – teaching and re – teaching and implementing various effective teaching methods (Maulana et al., 2015b, 2017). **The study developed by Opdenakker and Minnaert (2011) points out this ability of teachers to respond to students' different learning and basic psychological needs, as a critical factor of good teaching which may provide equal opportunities to all students regardless of their background characteristics.**

The mentioned six teaching **domains** can also be connected with Fuller's classical theory of teachers' stage concerns: learning climate is related with self – related concerns; classroom management and quality of instruction are associated with task- related concerns and the other three domains with students concerns (Fuller, 1969; Van de Grift et al., 2014). Besides, we can observe a cumulative order in the grades of complexity of these tasks so that those related with task concerns appear to be simpler than others like activating learning, teacher learning strategies and differentiation which involve an impact on students (Van de Grift et al., 2014).

Several studies have demonstrated that students' perceptions of their teachers' behaviour can predict their (self-reported) academic engagement, suggesting that the better the teaching behaviour perceived by students, the higher the level of academic engagement tends to be (Maulana et al., 2015a; Skinner & Belmont, 1993; Woolley & Bowen, 2007) or even the higher students' motivation is (Maulana et.al. 2015b). The importance of student engagement for various outcomes has been documented in the literature, including student learning and achievement, retention and graduation from secondary school, adjustment to school and admission and success in college (Finn, 1989; Fredricks et al., 2004; Opdenakker & Minnaert, 2011; Skinner & Belmont, 1993). Studies have also shown that student engagement can function as a protective factor for low achievement (Finn, 1993). Other researchers found rather large variations and instability in academic engagement over time and the importance of introducing new contents, student work time and closing components in teaching (Maulana, Opdenakker, Stroet & Bosker, 2012). Using an observation instrument to capture teachers' teaching behaviour and student engagement in the Netherlands, the study of Maulana et. al. (2017) shows that the six teaching behaviour domains are a reliable and valid measure of teaching behaviour, with a strong predictive validity for student engagement. They also found that two domains of

teaching behaviour - classroom management and clarity of instruction- appear to be more predictive of students' engagement compared to learning climate, activating learning, teaching learning strategies and differentiation, although these domains are important as well. Hence, student engagement can be seen as an indicator of good teaching behaviours and a mediator between classroom dynamics and student achievement (Furrer & Skinner, 2003; Virtanen, Lerkkanen, Poikkeus, & Kuorelahti, 2013).

Research also indicated that teaching behaviour follows a certain order in terms of level of complexity. Using a sample of pre-service teachers, Van de Grift et al. (2014) found the first three domains of teaching behaviour (learning climate, classroom management, clarity of instruction) are ordered as easier domains, while the other three (activating teaching, teaching learning strategies, differentiation) as more complex domains. These results were confirmed in the beginning teacher sample (Maulana et al., 2015) as well as in the more experienced teacher sample (Van der Lans, Van de Grift, & Van Veen, 2017). Research also suggests that teachers displaying more complex teaching skills, such as differentiation, have more positive influence on students' affective and cognitive outcomes (Creemers, Kyriakides, & Antoniou, 2009). Furthermore, there is also evidence that the quality of teaching behaviour depends on teaching experience. Van de Grift (2010) and Van de Grift, Van der Wal, and Torenbeek (2011) showed that the quality of teaching behaviour seems to be increasing as teachers gain experiences over time. The peak in the teaching behaviour quality seems to be visible when teachers reach 10 – 20 years of experience. Afterwards, its quality seems to decline when teachers become more senior professionals, towards the retirement period (Van de Grift et al., 2011).

Little is known whether the relationship between teaching behaviour and student engagement depends on teaching experience because research in this area is scarce. Insights from other relevant research suggest that teaching behaviours have important motivational effects, but the ways in which their behaviour affects students' engagement depends on students' general academic involvement and the importance that they attach to social relationships and emotional outcomes (Thijs & Verkuyten, 2010). Some studies have also analyzed the influence of teachers' teaching experience on teachers' sense of efficacy (Wolters & Daugherty, 2009). Sense of efficacy and teacher quality associated with the creation of caring and well-structured learning environment, with more positive teacher behaviors, attitudes, and interactions with students, mediate students' academic engagement (Rockoff, 2004; Klem & Connell, 2004). Because teaching quality seems to be connected to teaching experience, we assume that teaching experience will also play a role in explaining differences in the relationship between teaching behaviour and students' academic engagement.

To sum up, the decline of academic engagement can be connected with a decrease in the quality of teachers' teaching behaviours, and vice versa. Furthermore, effective teaching behaviour has a beneficial influence on engaging students academically, and the former can facilitate the achievement of higher grades and lower dropout rates. Based on the literature reviewed above, in the present study we sought to investigate the psychometric quality of Spanish version of the teaching behaviour instrument called My Teacher questionnaire (Maulana & Helms-Lorenz, 2016, 2017) for capturing student perceptions of teaching behaviour in the Spanish secondary education context. Furthermore, this study will contribute to validate the model of teaching behaviour and student engagement and its relevance for the Spanish context. Additionally, based on studies showing the relationship between teaching behaviour and teaching experience (Van de Grift, 2010; Van de Grift et al., 2011), we hypothesize that the relationship between teaching behaviour and student engagement will depend on teaching experience (differential effect).

3. Methodology

3.1. Participants

The participants were 7,114 students taught by 410 teachers attending 56 public and private schools in Spain. A total of 3,577 of the sample were boys (51%) and 3,415 were girls (49%). 122 students did not **disclose** their gender. Just under three quarters of the students (N = 5,112; 71.9 %) were in lower secondary education, 1,105 students (15.5%) were in upper secondary education and 897 students (12.6%) were in vocational education and training. A total of 3,183 students (44.7%) were at academic schools, 205 (2.9%) at vocational schools and 3,726 (52.4%) at schools which had academic and vocational programs simultaneously. A total of 4,702 students (66.1%) were at public schools whereas 2,412 (33.9%) were at private schools.

The initial intention of the research team was to use the probability proportional to size sampling technique. However, due to reticence from most schools we had to use a non-probabilistic convenience sampling method.

3.2. Measure

3.2.1. Teaching behaviour

To tap student perceptions of teachers' teaching behaviour, we used the My Teacher questionnaire based on the teaching behaviour model of Van de Grift (2007) and Van de Grift et al. (2014). The questionnaire was translated and back-translated for use in the Spanish context following the guidelines provided by Hambleton, Merenda, and Spielberger (2004). Two researchers with fluent English and deep knowledge of the Spanish education system conducted the initial Spanish translation. Subsequently, a university research panel assessed the translation results focusing on the item level to make sure that each item content was representative and relevant for the Spanish education system. Additionally, they gave opinions about the appropriate content and structure for use in the Spanish secondary education level. The initial Spanish translation was then translated back into English. The Spanish version and the back translated English version were checked by the second university research panel, including the original author of the questionnaire and a university professor of Spanish language.

The responses range from 1 (completely disagree) to 4 (completely agree). The total number of items is 41 divided into six **domains**: learning climate, classroom management, clarity of instruction, activating teaching, teaching learning strategies and differentiation (see Table 1).

3.2.2. Student engagement

To measure student engagement, the 10-items engagement scale of Skinner, Kindermann and Furrer (2009) was used. The scale consists of two **domains** of engagement (see Table 2): behavioural engagement (5 items) and emotional engagement (5 items). All responses were provided on a 4-point Likert scale, ranging from 1 (completely untrue) to 4 (completely true).

3.3. Procedure

In the spring term of the school year (March and April), the members of the research group requested student participation and collected data at each school. After a brief presentation in which the researchers described the purpose of the study, the students were asked to complete the questionnaire which took about 30 minutes. The questionnaires were distributed within normal class hours. There was no remuneration or course credit for participation and anonymity was guaranteed. No parents withheld their consent and all students accepted to cooperate answering the questionnaire.

3.4. Data analysis

Analyses were performed by dividing the sample into three subsamples. With the first subsample, an exploratory factor analysis (EFA) was carried out using the Factor program (Ferrando & Lorenzo-Seva, 2017). The second subsample was subject to a confirmatory factor analysis (CFA) with MPLUS 7.3 program (Muthén & Muthén, 2017). The third subsample was used for a second CFA, in order to confirm the previous CFA model.

We first checked whether data was suitable for carrying out EFA: normality of sample (skewness, kurtosis), the Bartlett's and the Kaiser-Meyer-Olkin's (KMO) indexes. Unweighted least squares were used as factor extraction method and the promim oblique was used as rotation method (Lorenzo-Seva, 2013). In line with the hypothesized model of teaching behaviour reviewed earlier (e.g., Van de Grift et al., 2014; Maulana et al., 2015a), we expect that six factors could be extracted. We chose the oblique method because the six factors of teaching behaviour are theoretically assumed to be correlated (Van de Grift, 2007; Maulana & Helms-Lorenz, 2016). Oblique approaches allow for the factors to be correlated (see Tabachnik & Fidell, 2007). Particularly, we opted for the promim oblique method which allows oblique rotations, but does not consider factors as pure measures of a single dimension.

The fitted CFA model included fit statistics: the Chi-Square test of significance (χ^2), the Tucker Lewis index–non normed fit index (TLI-NNFI), the Comparative Fit Index (CFI), the Goodness of Fit Index (GFI), Root Mean Square of Residuals (RMSR), Standardized Root Mean Square Residual (SRMR), and Steiger's Root Mean Square Error of Approximation (RMSEA). According to Hu and Bentler (1995) and Hooper et al. (2008), a good model fit, should maintain the following indexes: TLI >.90, CFI>.95, RMSEA y SRMR < de .08. Furthermore, we analyzed the multidimensional discrimination of items with MDISC index in order to test the quality of the measure and to use it as an indicator of the strength of the items within each domain. This index gave us the discrimination power of the item. The lower cut – off criteria is usually established in 0.2. Values showing more than 1.00 are considered as good discriminating items (Backer, 2001; Reckase, 2009; Ha, 2017).

Finally, multiple-group path analysis under the structural equation modelling (SEM) framework was conducted to test the relationship of the six teaching behaviour domains on students' behavioral and emotional engagement considering teachers' teaching experience.

4. Results

4.1. Teaching Behaviour

4.1.1. Exploratory factor analysis.

The proportion of variance explained by the six domains was 45%. Table 1 depicts the proportion of variance explained by each factor, together with the eigenvalues. The proportion of variance of the first factor was 27.19 and all the factors had eigenvalues larger than 1.00.

Insert Table 1

We found a Bartlett's statistic of = 28,576.9, $df = 820$, $p = .000010$, and KMO = 0.96. The scree plot (Cattell's test) was used as one of the indicators to determine how many factors were retained. The possible number of factors to retain was between two and six factors (Figure 1). We checked the communality of items and none of them showed less than 0.10 value of communality. The fit indices supported the six-factor solution as the best one compared with the two-factor solution, $\chi^2(2,329, 589) = 713.263$, $p = .000323$; TLI-NNFI = .999; CFI = .999; GFI = .996; RMSR = .019. Based on these results and combined with the original theoretical expectations regarding the six-domains of teaching behaviour, the six-factor solution was retained for further analyses.

The Cronbach's alpha coefficient for the whole six factors was 0.931. The Cronbach's alpha values for each factor were: learning climate = 0.66, efficient classroom management = 0.76, clarity of instruction = 0.70, activating teaching = 0.80, differentiation = 0.60 and teaching learning strategies = 0.71. Although there was an indication regarding the factor structure of teaching behaviour measure based on the EFA results, CFA was conducted to confirm the indicated factor structure. Additionally, a solution of two factors was also tested as informed by the Scree Plot. Although the comparative indexes were adequate, when considering the fit of absolute indexes the values were worse compared with the six-factor structure, $\chi^2(2,329, 701) = 3,391.065$, $p = .000010$; TLI-NNFI = .89; CFI = .90; GFI = .99; RMSR = .03). Furthermore, the percentage of explained variance was only 33%.

Insert Figure 1

Table 2 shows the values of MDISC index. Results show that all items had MDISC values above the cut-off criteria, which means that all items had sufficient discrimination index. The three items with highest multidimensional discrimination index were: “My teacher approaches me with respect (item 22)”, “My teacher motivates me to think (item 31)” and “My teacher asks me how I am going to learn the content of the lesson (item 16)”.

Insert Table 2

4.1.2. Confirmatory factor analysis

The first confirmatory factor analysis (CFA) which was conducted with the second subsample, $n = 2,380$, had the following fit indexes, TLI = .808, CFI = .821, RMSEA = .052, SRMR = .059. Although the TLI and CFI values were below the cut-off criteria of 0.90, the RMSEA and SRMR values were well below the cut-off criteria. This suggests that the model-data fit seems to be sufficient, but room for improvement is suggested. The model was replicated with the third subsample, $n = 2,405$, and the fit indexes were TLI = .820, CFI = .832, RMSEA = .043, SRMR = .058 (see Figure 2). Again with this subsample the model-data fit for the six factor solution seems to be acceptable.

Insert Figure 2

4.2. Academic engagement: behavioural engagement and emotional engagement.

4.2.1. Exploratory factor analysis.

The results showed two factors regarding academic engagement: behavioural engagement and emotional engagement. The Bartlett's statistic = 7,714, $df = 45$, $p = 0.000010$, and KMO = 0.87 indicated the adequacy of EFA, with the fit indexes: $\chi^2(2329, 26) = 147.029$, $p = 0.000010$; TLI-NNFI = 0.985; CFI = 0.992; GFI = 0.994; RMSR = 0.037. The alpha coefficient for the whole scale was 0.878. The multidimensional discrimination index (MDISC) showed that the ten items of the engagement measure had a good multidimensional discrimination (Table 3). The items which showed best discrimination between the possible answers were: “In this class, I pay attention” (item 4), “In this class, I listen very carefully” (item 5), and “In this class, it's fun” (item 8).

Insert Table 3

4.2.2. Confirmatory analysis

The analysis showed an acceptable fit with the second subsample, TLI = 0.882, CFI = 0.911, RMSEA = 0.093, SRMR = 0.051, and, with the third one, TLI = 0.877, CFI = 0.907, RMSEA = 0.095, SRMR = 0.054. The alpha coefficient for behavioural engagement was 0.93 and for emotional engagement 0.92.

Insert Figure 3

4.3. Teachers' teaching behavior, students' engagement, and teaching experience

Finally, a multiple-group path analysis under the SEM framework was conducted to see the influence of teaching behaviours on students' behavioral and emotional engagement. This analysis constitutes a first approximation, which needs to be deepened in future studies. Table 4 shows the relationship between teaching behaviour domains and student engagement. In general, the six teaching behaviour domains correlated more strongly with emotional engagement ($r = 0.24 - 0.31$) than with behavioural engagement ($r = 0.17 - 0.22$). Although small to moderate in magnitude, activating teaching seems to be the strongest predictor of both behavioural- and emotional engagement. The effect size for behavioural engagement range was between 3% and 5%. Activating teaching, efficient classroom management, and differentiation had effect sizes of 4.84% and 3.61% respectively. Regarding

students' emotional engagement, the effect sizes of the six domains had values between 6% and 10%. The highest effect sizes were found in activating teaching (9.61%), teaching learning strategies (7.29%), and differentiation (6.76%).

Insert Table 4

Furthermore, Figure 4 depicts the relational path between the six teaching behaviour domains and behavioral and emotional engagement, taking into account differences in teachers' teaching experiences. The categorization of teaching experience is based on Helms-Lorenz et al. (2018) and Van de Grift (2010) as follows: 0-2 years (beginner teachers), 3-9 years (less – experienced teachers), 10-19 years (moderately experienced teachers), 20-29 years (highly experienced teachers) and more than 30 years (senior experienced teachers).

The first model estimated the influence of the six-domains on students' emotional and behavioural engagement. However, results indicated that the model was just identified, which points to a zero degrees of freedom. This suggests that the model is not adequate for describing the relational path. Therefore, the subsequent models were modified based on the correlations between predictor variables and the criterion variables. Stepwise inclusion of predictors was done starting from the lowest correlate. The best model is represented by excluding the path from clarity of instruction to behavioural engagement, with the fit indices as follows: $\chi^2(7,092, 5) = 6.233, p = .2842$; TLI-NNFI = .99; CFI = 1.00; RMSEA = .01; SRMR = .004 (see Figure 4). Teaching learning strategies had significant and positive path to student behavioral engagement for beginner teachers (0-2 years of experience) and for highly experienced teachers (20-29 years of experience).

Regarding behavioral engagement, results showed that when teachers have more teaching experience, and the students perceive better teaching behaviour, students' behavioral engagement tends to be higher as well. For instance, the path for beginner teachers was only significant for teaching learning strategies ($\beta = .15; p < .05$), while the path for highly experienced and senior experienced teachers were significant for almost all domains, except efficient classroom management (where none of the teachers' experience categories revealed significant paths), and teaching learning strategies, where teaching learning strategies in senior experienced teachers revealed no significant effect on students' behavioral engagement.

For emotional engagement, results indicated that teachers with 10-19 years of experience (moderately experienced) showed significant paths in all domains, except in efficient classroom management, where no significant effects according to teachers' teaching experience were found. Teaching learning strategies seemed to depend on teachers' teaching experience in case of student emotional engagement: only beginner teachers showed a non-significant path, which might imply that having certain teaching experience is required to show a significant effect on students' emotional engagement.

Finally, it is worth mentioning that the percentage of explained variance in student behavioral engagement for the six teaching behaviour domains was as follows: beginner teachers (7%), less experienced teachers (7%), moderately experienced teachers (5%), highly experienced teachers (8%) and senior experienced teachers (6%). Regarding student emotional engagement, the explained variance for the six domains was: beginner teachers (15%), less experienced teachers (14%), moderately experienced teachers (13%), highly experienced teachers (14%) and senior experienced teachers (9%).

Insert Figure 4

5. Conclusions and Discussion

The present study investigates the psychometric quality of the teaching behaviour instrument called My Teacher questionnaire for capturing student perceptions of teaching behaviour in the Spanish secondary education context. Furthermore, this study aims to contribute to validate the model of

teaching behaviour and student engagement and its relevance for the Spanish context. Additionally, the study aims to provide the first attempt to explore the relationship between teaching behaviour and student engagement, taking into account differences in teaching experience (differential effect). Taken as a whole, the results of this research confirmed the factor structure of the original My Teacher and student engagement questionnaires

The questionnaire employed for this research consisted of two main constructs. The first one is teacher teaching behaviour measure, based on the model proposed by Van de Grift et al. (2014), and the second one is students' engagement based on the model proposed by Skinner et al. (2009). Our research confirms that the six teaching behaviour domains model is visible in the Spanish context as well. Furthermore, the MDISC value indicates an adequate functioning of all teaching behaviour items, which suggests that all items have a sufficient discrimination power. Regarding the reliability of each domain it is worth mentioning that although the differentiation domain showed a relatively low value (Cronbach's alpha = 0.60), which might well be due to too few items (4 items), the other domains have sufficiently high reliability values. Indeed the internal consistency of the whole scale was bigger than 0.90. Moreover, although most of the items have adequate values, the loadings of items 5 (0.37) and 10 (0.33) were rather low. Item 5 also has a relatively low loading in EFA (0.39, see Table 3). Typically, empirical research applies a cut-off point of 0.40. Because our goal is to have a construct that addresses many facets of a measured trait (i.e. teaching behaviour), we have retained those items as long as from the theoretical lens, they contribute to measure instructional clarity (item 5) and teaching learning strategies (item 10).

In line with other studies (e.g., Skinner et al., 2009; Maulana & Helms-Lorenz, 2017), results of the current study have confirmed the presence of two engagement domains (behavioural and emotional engagement). EFA showed a high construct consistency, and the reliability of each domain was above 0.90. As MSDISC index showed, the discrimination of items is also very good. The CFA results also revealed very high values with regard to the load of items. Only one item ("In the class I participate in class discussions") reached a relatively low value. In addition, its mean and MDISC values were lower compared with the rest of items. However, MDSIC value for this item was 0.35, which is higher than 0.20 (the cutoff value for this index). This item also has a relatively low loading in EFA. However, when this item was deleted, the internal consistency and the alpha values decreased. Hence, we decided to retain this item as an indicator of behavioural engagement.

This study supports the assumption that teachers' teaching behaviour in Spain can be studied in terms of the six domains as well, including learning climate, efficient classroom management, clarity of instruction, activating teaching, differentiation, and teaching learning strategies. Furthermore, the correlations with student academic engagement revealed that teachers' behaviours have sufficient predictive power. Results indicate that teaching behaviours appear to be better predictors of students' emotional engagement compared with behavioural engagement. Nevertheless, the predictive value for behavioural engagement remains important as well. These findings are consistent with other research, which provide empirical evidence for the link between teaching behaviour and students' academic engagement (e.g., Davidson, Gest, & Welsh, 2010; Furrer & Skinner, 2003; Maulana et al., 2015a, 2017; Opdenakker, Maulana, & Den Brok, 2012). The strongest relationship is between activating teaching and emotional engagement. This finding is expected given that the content of this domain refers to teachers' behaviour focusing on students' feelings and motivation. This finding is consistent with another study associated with the predictive quality of student perceptions on student outcomes (Maulana & Helms-Lorenz, 2016).

Differentiation and teaching learning strategies were the second and the third strongest in terms of their relationship with emotional engagement. In the context of other studies addressing the relation between teaching factors and students outcomes and engagement (Furrer & Skinner, 2003; Hattie, 2009, 2012; Kyriakides et al., 2013), our results point to the importance to consider further analysis to determine if teachers' skills and instructional strategies could improve their students' emotional engagement. On the other hand, students' behavioural engagement shows a good

correlation with teachers' efficient classroom management, their skill to develop activating teaching, and differentiation.

Finally, the present study shows that although teaching behaviour domains are generally important for students' academic engagement, the relationship between teaching behaviour and engagement seems to depend on teachers' teaching experience. In general, findings highlighted that emotional engagement seems to be more strongly related to student perceptions of teaching behaviour than behavioral engagement. The percentage of explained variance is bigger for emotional engagement than for behavioral engagement. Nevertheless, this influence tends to decrease as teachers' teaching experience increases, except for the case of very experienced teachers who show and slight improvement. Furthermore, the efficient classroom management domain should be studied further in future research, especially because it only has a significant path on students' behavioral engagement in the case of moderately experienced teacher (10-19 years of experience).

In conclusion, the current study indicates that the psychometric quality of My Teacher questionnaire for capturing student perceptions of teaching behaviour, in the Spanish secondary education context is adequate. The relevance of teaching behaviour and student engagement for the Spanish context is evident. Teaching experience seems to influence the relationship between teaching behaviour and student engagement. This study is highly relevant particularly for many Spanish-speaking contexts worldwide, but also for other contexts interested in teaching behaviour comparison from the student perspective. The instrument is also useful for improving teacher practice by using it as a tool for teacher professional development. The instrument might be useful for low-stake as well as high-stake evaluations in many Spanish-speaking countries, but cautions should be made when using the instrument for those purposes (Van der Lans & Maulana, 2018).

6. Limitations

Although this study yields important implications for research and educational practice, it also has several limitations. Firstly, although the sample is quite large, it does not cover all the regions of Spain. Therefore, interpretation of findings regarding all Spanish population should be handled with caution until research involving more representative sample is available. Hence, future research should try to enlarge the sample throughout the country to validate the more actual factorial structure. This strategy could improve the generalization of the current factorial structure to the whole country. It would also be interesting to focus on a comparative and international perspective to share results and facilitate the analysis of differences around the world. The students participated in this study on a voluntary basis. Future research should attempt to increase the number of participants and randomly sample them, if possible.

Furthermore, running a multilevel analysis for hierarchically structured data such as ours should be done in the future whenever possible. It is also important to deepen our understanding of teaching effective behaviours because this knowledge may give us clues about how to adapt initial and continuous teachers' training to their real and actual needs. This knowledge will offer clues about how teachers should pay attention to particular students in order to encourage a more personal learning with better results. Some studies (Furrer & Skinner, 2003; Lietart, Roorda, Laevers, Verschueren, & De Fraine, 2015) found interesting differences in students' engagement according to their age and gender. More studies in the Spanish context are needed to test whether those student factors also matter for their learning engagement. The present study focuses on the relationship between teachers' behaviours and students' engagement by treating teaching behaviour domains as predictors and student engagement as an outcome measure (uni-directional). We expect that the relationship between these two constructs might be dynamic and bi-directional. The body of knowledge would benefit from studies addressing the opposite direction and the reciprocal relationship between these two constructs. Additionally, previous research indicates that regarding academic engagement and motivation, student perceptions of teaching behaviour are more predictive than observation (De Jong & Westerhof, 2001; Maulana & Helms-Lorenz, 2016). This study lacked other measures such as information from other agents. Moreover, the measurement of teaching

behaviour should also include complementary sources of information such as teachers' self report and school principals' opinions in order to assess external validity and to triangulate different sources of information.

Finally, longitudinal studies are needed in which students are followed during several years in their lower and upper secondary education. This would allow the investigation of the changes in student engagement and also in relation to their teachers' behaviours across secondary education.

7. References

- Anderson, A. R., Christenson, S. L., Sinclair, M. F., & Lehr, C.A. (2004). Check and connect: the importance of relationships for promoting engagement with school. *Journal of School Psychology, 42*, 95-113. doi: 10.1016/j.jsp.2004.01.002
- Aaronson, D., Barrow, L., & Sander, W. (2007). Teachers and Student achievement in the Chicago Public High Schools. *Journal of Labor Economics, 25*(1) 95-135.
- Antoniou, P., A., Kyriakides, L., & Creemers, B. (2011). Investigating the effectiveness of a dynamic integrated approach to teacher professional development. *CEPS Journal, 1*, 13-41.
- Baker, F. (2001). *The Basics of Item Response Theory* (2nded.). Madison: ERIC Clearinghouse on Assessment and Evaluation. Retrieved from: <https://files.eric.ed.gov/fulltext/ED458219.pdf>
- Bosker, R. J. & Witziers, B. (1996). *The magnitude of school effects, or: Does it really matter which school a student attends?* Paper presented at the annual meeting of the American Educational Research Association, New York.
- Burdsal, C. A. & Bardo, J. W. (1986). Measuring student's perceptions of teaching: dimensions of evaluation. *Educational and Psychological Measurement, 46*, 63-79.
- Centra, J. A. & Potter, D. A. (1980). School and teacher effects: an interrelational model. *Review of Educational research, 50*(2), 273-291.
- Creemers, B. P. M. & Kyriakides, L. (2008). *The dynamics of educational effectiveness: a contribution to policy, practice and theory in contemporary schools*. London and New York: Routledge.
- Danielson, C. (2013). *Rubrics form the framework for teaching evaluation instrument*. Retrieved from: <https://www.danielsongroup.org/framework/>
- Davidson, A. L., Gest, S. D., & Welsh, J. A. (2010). Relatedness with teachers and peers during early adolescence: an integrated variable – oriented and person – oriented approach. *Journal of School Psychology, 48*, 483-510. doi: 10.1016/j.jsp.2010.08.002
- De Jong, R. & Westerhof, K. J. (2001). The quality of student ratings of teacher behaviour. *Learning Environments Research, 4*, 51-85. doi: 10.1023/A:1011402608575
- Ferrando, P. J. & Lorenzo-Seva, U. (2017). Program Factor at 10: Origins, development and future directions. *Psicothema, 29*(2), 236-240. doi: 10.7334/psicothema2016.304.
- Finn, J. D. (1989). Withdrawing from school. *Review of Educational Research, 59*, 117 - 142.
- Finn, J. D. (1993). *School engagement and students at risk*. Washington: National Center for Educational Statistics.

- Fredricks, J., McColskey, W., Meli, J., Mordica, J., Montrosse, B., & Mooney, K. (2011). *Measuring student engagement in upper elementary through high school: a description of 21 instruments.* (Issues & Answers Report, REL 2011–No. 098). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southeast. Retrieved from: https://ies.ed.gov/ncee/edlabs/regions/southeast/pdf/REL_2011098.pdf
- Fuller, F. (1969). Concerns of teachers: a developmental conceptualization. *American Educational Research Journal*, 6(2), 207-226.
- Furrer, C. & Skinner, E. (2003). Sense of relatedness as a factor in children's academic engagement and performance. *Journal of Educational Psychology*, 95(1), 148-162. doi: 10.1037/0022-0663.95.1.148.
- Guldmond, H. & Bosker, R. J. (2009). School effects on students' progress – a dynamic perspective. *School Effectiveness and School Improvement*, 20(2), 255-268. doi: 10.1080/09243450902883938
- Guskey, T. R. & Passaro, P. D. (1994). Teacher efficacy: a study construct dimensions. *American Educational Research Journal*, 31(3), 627-643.
- Ha, T. D. (2017). Applying Multidimensional Three-Parameter Logistic Model (M3PL) in Validating a Multiple-Choice Test. *International Journal of Scientific and Research Publications*, 7(2), 175-183.
- Hambleton, R. K., Merenda, P., & Spielberger, C. (Eds.). (2004). *Adapting educational and psychological tests for cross-cultural assessment*. Hillsdale, NJ: Lawrence Erlbaum Publishers.
- Hattie, J. A. C. (2003, October). Teachers make a difference: What is the research evidence? Paper presented at the Building Teacher Quality: What does the research tell us ACER Research Conference, Melbourne, Australia. Retrieved from http://research.acer.edu.au/research_conference_2003/4/
- Hattie, J. A. C. (2009). *Visible learning: A synthesis of over 800 meta – analyses relating to achievement*. New York: Routledge.
- Hattie, J. A. C. (2012). *Visible learning: maximizing impact on learning*. Abingdon: Routledge.
- Hooper, D., Coughlan, J., & Mullen, M. (2008). Structural Equation Modelling: Guidelines for Determining Model Fit. *The Electronic Journal of Business Research Methods*, 6(1), 53-60. Retrieved from www.ejbrm.com.
- Houtveen, A. A. M., Van de Grift, W., & Brokamp, S. K. (2014). Fluent Reading in Special Elementary Education. *School Effectiveness and School Improvement*, 25(4) 555-569.
- Helms-Lorenz, M., Maulana, R., Teli, S., Inda – Caro, M., Fernandez – Garcia, C-M., & Attard Tona, M. (2018, January). *Measuring effective teaching behaviour using student questionnaire: Multinational perspective*. Paper presented during the annual International Congress for School Effectiveness and Improvement, Singapore.
- Hoyt, D. P. & Pallet, W. H. (1999). Appraising teaching effectiveness: beyond student ratings. *Idea Paper*, 36, 1-14.
- Hu, L. T. & Bentler, P. (1995). Evaluating Model Fit. In R. Hoyle (ed.), *Structural Equation Modelling. Concepts, Issues, and Application* (pp. 76-99). London: SAGE.

- Klem, A. M. & Connell, J. P. (2004). Relationships matter: linking teacher support to student engagement and achievement. *Journal of School Health*, 74(7), 262-273.
- Konstantopoulos, S. & Sun, M. (2014). Are teacher effects larger in small classes? *School Effectiveness and School Improvement*, 25(3), 312-328. doi: 10.1080/09243453.2013.808233
- Kyriakides, L., Christoforou, C., & Charalambous, C. Y. (2013). What matters for student learning outcomes: a meta – analysis of studies exploring factors of effective teaching. *Teaching and Teacher Education*, 36, 143-152. doi: 10.1016/j.tate.2013.07.010
- Kyriakides, L. & Creemers, B. P. M. (2009). The effects of teacher factors on different outcomes: two studies testing the validity of the dynamic model. *Effective Teaching*, 1(1), 61-85. doi: 10.1080/19415530903043680
- Kyriakides, L., Creemers, B. P. M., & Antoniou, P. (2009). Teacher behaviour and student outcomes: suggestions for research on teacher training and professional development. *Teaching and Teacher Education*, 25(1), 12-23. doi: 10.1016/j.tate.2008.06.001
- Lietaert, S., Roorda, D., Laevers, F., Verschueren, K., & De Fraine, B. (2015). The gender gap in student engagement: the role of teachers' autonomy support, structure and involvement. *British Journal of Educational Psychology*, 85, 498-518. doi: 10.1111/bjep.12095.
- Lorenzo-Seva, U. (2013). Why rotate my data using Promin? Technical Report. Retrieved from <http://psico.fcep.urv.es/utilitats/factor/documentation/whypromin.pdf>.
- Maulana, R. & Helms-Lorenz, M. (2016). Observations and student perceptions of the quality of preservice teachers' teaching behaviour: construct representation and predictive quality. *Learning Environment Research*, 19, 335-357. doi: 10.1007/S10984-016-9215-8
- Maulana, R. & Helms-Lorenz, M. (2017, April). Student engagement and beginning teacher differentiation practices in Dutch secondary education: Are there gender differences? *Poster presented at the Annual Meeting of American Educational Research Association*, San Antonio, TX.
- Maulana, R. & Schuurman, G. (2018, February). *ICALT observation training*. Power point presented during the ICALT training in Oviedo, University of Oviedo, Spain.
- Maulana, R., Helms-Lorenz, M., & Van de Grift, W. (2015a). Development and evaluation of a questionnaire measuring pre – service teachers' teaching behaviour: a Rasch modelling approach. *School effectiveness and school improvement; an International Journal of Research, Policy and Practice*. 26(2), 169-194. doi: 10.1080/09243453.2014.939198
- Maulana, R., Helms-Lorenz, M., & Van de Grift, W. (2015b). Pupils' perception of teaching behaviour: evaluation of an instrument and importance of academic motivation in Indonesian secondary education. *International Journal of Educational Research*, 69, 98-112. doi: 10.1016/j.ijer.2014.11.002
- Maulana, R., Helms-Lorenz, M., & Van de Grift, W. (2017). Validating a model of effective teaching behaviour of pre – service teachers. *Teachers and teaching*, 23(4), 471-493. doi: 10.1080/13540602.2016.1211102
- Maulana, R., Opdenakker, M. C., Stroet, K., & Bosker, R. (2012). Observed lesson structure during the first year of Secondary Education: exploration of change and link with academic engagement. *Teaching and Teacher Education*, 28, 835-850. doi: 10.1016/j.tate.2012.03.005.

- Muijs, D., Kyriakides, L., Van der Werf, G., Creemers, B. P. M., Timperley, H., & Earl, L. (2014). State of the art – teacher effectiveness and professional learning. *School Effectiveness and School Improvement*, 25(2), 231-256. doi: 10.1080/09243453.2014.885451
- Muijs, D., Campbell, J., Kyriakides, L., & Robinson, W. (2005). Making a case for differentiated teacher effectiveness: an overview of research in four key areas. *School Effectiveness and School Improvement*, 16(1), 51-70. doi: 10.1080/09243450500113985
- Muthén, L. & Muthén, B. (2014). *Mplus. Statistical Analysis*. Los Ángeles: Muthén & Muthén.
- Newmark, D. (1929). Students' opinions of their best and poorest teachers. *The Elementary School Journal*, 29(8), 576-585.
- Opdenakker, M. C., Maulana, R., & Den Brok, P. J. (2012). Teacher – student interpersonal relationships and academic motivation within one school year: developmental changes and linkage. *School effectiveness and school improvement; an International Journal of Research, Policy and Practice*, 23(1), 95-119. doi: 10.1080/09243453.2011.619198
- Opdenakker, M. V. & Minnaert, A. (2011). Relationships between learning environment characteristics and academic engagement. *Psychological Reports*, 109(1), 259-284. doi: 10.2466/09.10.11.PR0.109.4.259-284
- Opdenakker, M. C. & Van Damme, J. V. (2001). Relationships between school composition and characteristics of school processes and their effects on mathematics achievement. *British Educational Research Journal*, 27(4), 407-432. doi: 10.1080/01411920120071434
- Opdenakker, M. C. & Van Damme, J. V. (2006). Differences between Secondary Schools: a study about school context, group composition, school practice, and school effects with special attention to public and catholic schools and types of schools. *School Effectiveness and School Improvement*, 17(1), 87-117. doi: 10.1080/09243450500264457
- Opdenakker, M. C., Van Damme, J. V., De Fraine, B., Van Landeghem, G., & Onghena, P. (2002). The effect of schools and classes on mathematic achievement. *School Effectiveness and School Improvement*, 13(4), 399-427.
- Pianta, R. C. & Hamre, B. K. (2009). Conceptualization, measurement and improvement of classroom processes: standardized observation can leverage capacity. *Educational Researcher*, 38(2), 109-119. doi: 10.3102/0013189X09332374
- Reckase, M. (2009). Statistical Descriptions of Item and Test Functioning. In S. E. Fienberg & W. J. van der Linden (Eds.), *Multidimensional Item Response Theory* (pp. 113-136). Michigan: Springer.
- Rockoff, J. E. (1994). The Impact of Individual Teachers on Student Achievement: Evidence from Panel Data. *The American Economic Review*, 94(2), 247-252. doi: 10.1257/0002828041302244
- Scheerens, J. (2016). Theories on educational effectiveness and ineffectiveness. In *Educational Effectiveness and Ineffectiveness* (pp. 259-289). Springer: The Netherlands.
- Seidel, T., & Shavelson, R. J. (2007). Teaching effectiveness research in the past decade: The role of theory and research design in disentangling meta-analysis results. *Review of Educational Research*, 77(4), 454-499. doi: 10.3102/0034654307310317.

- Skinner, E. A. & Belmont, M. (1993). Motivation in the classroom: reciprocal effects of teacher behavior and student engagement across the school year. *Journal of Educational Psychology*, 85(4), 571-581. doi: 10.1037/0022-0663.85.4.571.
- Skinner, E. A., Kindermann, T. A., & Furrer, C. J. (2009). A Motivational Perspective on Engagement and Disaffection Conceptualization and Assessment of Children's Behavioral and Emotional Participation in Academic Activities in the Classroom. *Educational and Psychological Measurement*, 69(3), 493-525. doi: 10.1177/0013164408323233
- Snijders, T., & Bosker, R. (2012). *Multilevel analysis* (2nded.). London: SAGE.
- Staiger, D. O. & Rockoff, J. E. (2010). Searching for effective teachers with imperfect information. *Journal of Economic Perspectives*, 24(3), 97-118. doi: 10.1257/jep.24.3.97.
- Tabachnick, B. G. & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). Upper Saddle River, NJ: Pearson Allyn & Bacon
- Teodorovic, J. (2011). Classroom and school factors related to student achievement: what works for students? *School Effectiveness and School Improvement*, 22(2), 215-236. doi: 10.1080/09243453.2011.575650.
- Thijs, J. & Verkuyten, M. (2010) Students' Anticipated Situational Engagement: The Roles of Teacher Behavior, Personal Engagement, and Gender. *The Journal of Genetic Psychology*, 170(3), 268-286. doi: 10.1080/00221320903218323
- Townsend, T. (2007). *International Handbook of School Effectiveness and Improvement*. Dordrecht, the Netherlands: Springer.
- Van den Broeck, A., Opdenakker, M. C., & Van Damme, J. (2005). The effects of student characteristics on Mathematics achievement in Flemish TIMSS 1999 data. *Educational Research and Evaluation*, 11(2), 107-121. doi: 10.1080/13803610500110745
- Van de Grift, W. (2007). Quality of teaching in four European countries: a review of the literature and application of an assessment instrument. *Educational Research*, 49(2), 127-152. doi: 10.1080/00131880701369651
- Van de Grift, W. (2010). *Ontwikkeling in de beroepsvaardigheden van leraren* [Development of teachers' professional skills]. Groningen: Rijksuniversiteit Groningen
- Van de Grift, W. (2014). Measuring teaching quality in several European countries. *School effectiveness and school improvement. An International Journal of Research, Policy and Practice*, 25(3), 295-311. doi: 10.1080/09243453.2013.794845
- Van de Grift, W., Helms-Lorenz, M., & Maulana, R. (2014). Teaching skills of student teachers: calibration of an evaluation instrument and its value in predicting student academic engagement. *Studies in Educational Evaluation*, 43, 150-159. doi: 10.1016/j.stueduc.2014.09.003
- Van der Lans, R., & Maulana, R. (2018). The use of secondary school student ratings of their teacher's skillfulness for low-stake assessment and high-stake evaluation. *Studies in Educational Evaluation*, 58, 112-121. doi: 10.1016/j.stueduc.2018.06.003
- Van der Lans, R. M., Van de Grift, W., & Van Veen, K. (2015). Developing a teacher evaluation instrument to provide formative feedback using student ratings of teaching acts. *Educational Measurement: issues and practices*, 34(3), 18-27. doi: 10.1111/emip.12078

- Van de Grift, W., Van der Wal, M., & Torenbeek, M. (2011). Ontwikkeling in de pedagogische didactische vaardigheid van leraren in het basisonderwijs. [The development of pedagogical didactical competencies of primary school teachers]. *Pedagogische Studiën*, 88(6), 416-432.
- Van de Grift, W. & Van Veen, K. (2017). Individual differences in teacher development: An exploration of the applicability of a stage model to assess individual teachers. *Learning and Individual Differences*, 58, 46-55. doi: 10.1016/j.lindif.2017.07.007
- Virtanen, T. E., Lerkkanen, M. J., Poikkeus, A. M., & Kuorelahti, M. (2013). The relationship between classroom quality and students' engagement in secondary school. *Educational Psychology*, 35(8), 963-983. doi: 10.1080/01443410.2013.822961
- Wolters, C. A. & Daugherty, S. G. (2009). Goal structures and teachers' sense of efficacy: their relation and association to teaching experience and academic level. *Journal of Educational Psychology*, 99(1), 181-193. doi: 10.1037/0022-0663.99.1.181
- Woolley, M. E. & Bowen, G. (2007). In the context of risk: supportive adults and the school engagement of middle school students. *Family relations*, 56(1), 92-104. doi: 10.1111/j.1741-3729.2007.00442.x