Culture and Education

Teaching behaviours in Spain under observation: an instrument for assessing teaching quality

--Manuscript Draft--

Full Title:	Teaching behaviours in Spain under observation: an instrument for assessing teaching quality
Manuscript Number:	RCYE-2021-0021R2
Order of Authors:	Carmen-María Fernández - García
	Mercedes Inda-Caro
	Ridwan Maulana
	Susana Torío - López
Article Type:	Research Paper
Keywords:	observation; teaching quality; teaching effectiveness; Assessment
Abstract:	Teachers constitute a key point in the educational process. Therefore having information about their behaviours inside the classroom can give us very useful data when the intention is to improve teaching effectiveness. This study aims to describe teaching behaviours observed in 344 teachers in 56 public and private schools in Asturias (Spain). Descriptive analyses and analysis of variance have been run out in order to answer the research questions. A stepwise regression has been performed too, to identify which teaching skill domains are more important in promoting “student engagement”. Results have shown positive, significant relationships between all the domains and “student academic engagement”. “Activating teaching”, “efficient classroom management” and “teaching-learning strategies” seem to be the main teaching skills for increasing “student engagement”. Interesting differences have also been found in “student engagement” depending on the standard of teachers’ teaching behaviour.
Response to Reviewers:	Reviewer 2: Please add comments you don't mind the author seeing. (Por favor agregar comentarios no le importa el autor ver.) © The only important point to improve is to explain the practical use of the instrument. We have improved the explanation of the practical use of the instrument: oriented to the analysis of verbal interactions and to the reflection about teachers’ own practices. The process of evaluation of teaching methods and the whole process of interaction should be understood and assessed in more accesible qualitative manner. Please, try to relate your article more to historical cultural conception. We have included these references to reinforce the theoretical framework concerning the six ICALT domains. We have also emphasized the need to put previous experiences ‘in context’ and to consider the importance that social interaction has for learning. The historical cultural conception has also been considered in the conclusions and limitations sections. The explanation of the use of the instrument is still needed together with its justification from the point of view of historical cultural approach. We have included this justification: This instrument also gives an interesting opportunity to analyse verbal interactions which represent teachers and students’ experiences and which constitute a key point to understand all the personal transformations which take place in the process of teaching and learning. It can also provide a framework to reflect about their own practice, how do they cope with complex decisions regarding their teaching and how can they improve this process (Pérez, 2013). There are no references of the authors of cultural historical approach for humano development or learning process. We advise the authors to include precise citations of historical cultural approach for teaching and learning. We have included these new references which directly or indirectly consider the historical cultural approach for teaching and learning: Carriedo, N. (1995). Hacia la contextualización: la enseñanza de estrategias de

comprensión de las ideas principales en el aula. Comunicación, Lenguaje y Educación, 28, 123-134.
Cole, M. (1984). La zona de desarrollo próximo: donde cultura y conocimiento se generan mutuamente. Infancia y Aprendizaje, 25, 3-17.
Coll, C. (1984). Estructura grupal, interacción entre alumnos y aprendizaje escolar. Infancia y Aprendizaje, 27/28, 119-138.
Coll, C. (1988). Significado y sentido en el aprendizaje escolar. Reflexiones en torno al concepto de aprendizaje significativo. Infancia y Aprendizaje, 41, 131-142.
Cubero, R., Cubero, M., Santamaría, A., Saavedra, J. & Yossef, J.J. (2007). Aprendizaje y psicología histórico-cultural. Aportaciones de una perspectiva social del aula. Investigación en la escuela, 62, 5-16. https://doi.org/10.12795/IE.2007.i62.01
Gispert, I. & Onrubia, J. (1997). Analizando la práctica educativa con herramientas socio-culturales: traspaso del control y aprendizaje en situaciones de aula. Cultura y Educación, 9(2/3), 105-115. https://doi.org/10.1174/113564097761403526
Goodnow, J.J. (1996). Acceptable ignorance, negotiable disagreement: alternative reviews of learning. In D.R. Olson & N. Torrance (Eds.), The handbook of education and human development (pp. 345-367). Blackwell Publishers.
John –Steiner, V. & Mahn, H. (1996). Sociocultural approaches to learning and development: a Vygostkyan framework. Educational Psychologist, 31(3/4), 191-206.
Mercado, R. (1991). Los saberes docentes en el trabajo cotidiano de los maestros. Infancia y Aprendizaje, 55, 59-72.
Pérez, M. (2013). La actitud indagadora del profesor: un proceso para desarrollar competencias en los docentes. Profesorado. Revista de curriculum y formación del profesorado, 17(3), 57-72.
Prado, M. & Cubero, M. (2013). Reflexionando acerca de cómo estudiar el proceso de enseñanza-aprendizaje en el contexto natural del aula universitaria. Cultura y Educación, 25(3), 273-284. https://doi.org/10.1174/113564013807749696
Vygostky, L. S. (1979). Pensamiento y lenguaje. La Pléyade.
Vygostky, L. S. (1984). Aprendizaje y desarrollo intelectual en la edad escolar. Infancia y Aprendizaje, 27/28, 105-116.

Funding Information:	Dutch scientific funding agency/INIE (NRO, under grant 405-15-732; INIE-19-MOD C-1)	Not applicable
Additional Information:		
Question	**Response**	

Author identification Data

Carmen- María Fernández – García. Facultad Formación Profesorado y Educación. Universidad de Oviedo (España) C/ Aniceto Sela s/n. Room 213 (33005) Oviedo. Asturias. +34 985103271. fernandezcarmen@uniovi.es

Mercedes Inda – Caro. Facultad Formación Profesorado y Educación. Universidad de Oviedo (España) C/ Aniceto Sela s/n. Room 314 (33005) Oviedo. Asturias. +34 985102857. indamaria@uniovi.es

Ridwan Maulana. Faculty of Behavioural and Social Sciences. Universidad de Groningen (Holanda) Grote Kruisstraat 2/1, (9712TS), Groningen. r.maulana@rug.nl

Susana Torío – Lopez. Facultad Formación Profesorado y Educación. Universidad de Oviedo (España). C/ Aniceto Sela s/n. Room 249 (33005) Oviedo. Asturias. +34 985103244. storio@uniovi.es

Corresponding Author:

Carmen- María Fernandez – García. fernandezcarmen@uniovi.es. Facultad de Formación del Profesorado y Educación, C/ Aniceto Sela s/n. Room 213. (33005) Oviedo, Asturias.

Teaching behaviours in Spain under observation: an instrument for assessing teaching quality

La observación del comportamiento del profesorado: un instrumento para evaluar la calidad docente

7973 words

Acknowledgements

This work was supported by the Dutch scientific funding agency (NRO, under grant 405-15-732) and the Institute of Educational Research and Innovation of the University of Oviedo (INIE, under grant INIE-19- MOD C-1).

Abstract

Teaching quality affects student outcomes and is also reflected in the results obtained in international tests. Although several factors determine the results obtained by students in the already mentioned tests and their general academic performance, undoubtedly teachers constitute a key point. This study aims to describe teaching behaviours observed in 344 teachers in 56 public and private schools in Asturias (Spain). Descriptive analyses and analysis of variance have been run out in order to answer the research questions. A stepwise regression has been performed too, to identify which teaching skill domains are more important in promoting student engagement. Results have shown positive, significant relationships between all the domains and student academic engagement. Activating teaching, efficient classroom management and teaching and learning strategies seem to be the main teaching skills for increasing student engagement. Interesting differences have also been found in student engagement depending on the standard of teachers' teaching behavior.

Key words: observation; teaching quality; teaching effectiveness; assessment

Student learning depends on several factors which relate to students' own capabilities and motivation but also to their family background and teachers' skills. Undoubtedly, teachers are a key point in the learning process (Hanushek, 2016), and several theoretical models and authors (Danielson, 1996; De Jong & Westerhoff, 2001; Fuller, 1969; Hattie, 2003; Kyriakides et al., 2009; Kugel, 1993; Muijs & Reynolds, 2000; Pianta & Hamre, 2009; Sammons & Bakkun, 2011; Van de Grift, 2007) have explained the development of teachers' practices trying to identify which skills help teachers to be more effective and obtain better outcomes from their students.

Measuring teaching effectiveness

We have evidence (Briole & Maurin, 2019) that evaluating teachers generates significant benefits not only for teachers (e.g. improvement of their core skills) but also for students (e.g. improvement of educational equality).Theories of teacher development have been studied widely, resulting in different instruments for assessing teaching quality. Danielson (1996) has developed a model to assess effective teaching which has been divided into 22 components clustered in four domains of teaching responsibility: planning and preparation, classroom environment, instruction and professional responsibilities. In 2007 she developed a new framework which has resulted in a rubric from the Framework for Teaching Evaluation instrument (Danielson, 2013), including clear standards of practice.

Hamre and Pianta have also presented a scientific framework to assess classroom quality, the Classroom Assessment Scoring System (CLASS). This framework has assessed three domains of quality (emotional support, instructional support, and classroom organization) considering teacher–student interactions as being likely to contribute positively to students' development as a consequence of their experience in the classroom (Pianta & Hamre, 2009). In this same line, the teaching through interaction framework, has improved the CLASS system but considering the same three core domains (Hamre et al., 2013).

Also in the United States, Reddy & Dudek (2014), have stablished a system for teaching assessment including instructional and behavioral management strategies. The model, which has been based in the exhaustive revision of previous assessment models has considered whether strategies are used for individual students or groups, the inclusion of summaries of concepts, the presence of corrective feedback, the promotion of direct instruction, adaptive instruction or student thinking, to mention some examples (Reddy et al., 2013). This system has combined external observations of teachers and self-evaluation (Reddy et al., 2013) with the intention to promote a dialogue between teachers and observers.

In the European context (Cyprus), research during the 1990s can be seen as the starting point for a dynamic model which has tried to describe the complex nature of educational effectiveness. The model has considered multiple factors which have an influence at different levels: student, teacher/classroom, school and educational system. Focusing on teachers, the model has set out observable instructional behaviours grouped into eight factors: orientation, structuring, questioning, teaching modelling, application, time management, teacher role in making the classroom a learning environment and classroom assessment (Kyriakides et al., 2009).

The education inspectorates in several European regions have also undertaken comparative analyses of effective teaching using the observation instrument developed by Van de Grift and Lam (1998), which is a reliable, valid way to compare educational data, regardless of cultural differences between countries. The International Comparative Analysis of Learning and Teaching Project (ICALT) has integrated this instrument and the state of the art on teaching effectiveness in its procedure, also including non-European countries and secondary education. The adaptation of the initial domains has resulted in a final structure of six domains: safe and stimulating learning climate, classroom management, clarity of instruction, activating teaching, teaching and learning strategies and differentiation. There is a graduation

of levels of complexity in these domains. All of them have impact on students but the most complex ones are not easily acquired or deployed by most teachers (authors, 2014). In spite of this, teaching development should not be seen as a succession of rigid stages: Van der Lans et al. (2015) have demonstrated that the least complex skills in more complex domains may precede the development of the most complex skills in other less complex domains. The six domains are discussed in more detail below.

A safe and stimulating learning climate is one of the core domains because of its influence on student learning results and engagement (Reyes et al., 2012). Moreover, this influence of school climate have determined 20-40% of student achievement according to Van de Grift (2007). Although consensus about the exact characteristics of this kind of learning climate has not been reached, Wang et al., (1993) have reinforced the importance of this idea when they conclude that variables that affect students directly are the most determinant ones. Therefore, other domains such as those included in Danielson's (1996) model - school and policy level- despite needing to be considered, do not seem to be so determinant. This domain has a clear connection with teachers and students interactions, which have been identified as one of the most important aspects of teachers' job and so determinant to understand students' achievement, attitudes and motivation (Hamre et al., 2013; Howe et al., 2019).

Efficient classroom management allows teachers to achieve their objectives more easily, as it includes time management, orderly presentation of content, ensuring that the lesson begins and ends on time, appropriate balance of individual/group activities and effectively dealing with student misbehavior (Danielson, 1996; Oliver & Reschly, 2007; Van de Grift, 2007). Emmer & Stough (2001) have also demonstrated that efficient classroom management can reduce behaviour and discipline problems, leaving more time for educational purposes and providing more opportunities to learn. Nevertheless, other researchers have related classroom management practices with student engagement but not with disruptive behaviours (Gage et

al., 2018). On similar lines, research has suggested that effective teachers spend 15% less time on management and organization and 50% more time on instruction and interactive activities (Van de Grift, 2007), so effectively managing time leaves more time for direct educational purposes.

Clarity of instruction entails aspects of instructional quality such as giving staged instructions, making clear whether an answer is right or wrong and regularly checking if learners have understood what the lesson is about. Kyriakides et al. (2009) have referred to this teachers' intention to order and organize lesson information, as structure. Hence, students will not learn as much as they could, if instructions are unclear (Authors, 2017b). This domain has been connected with the idea of significant learning in which students' prior knowledge must be considered, giving them the opportunity to meaningfully learn in their lessons. Additionally, it requires explicit objectives, information about mistakes and the procedures required during each class (Blaich et al., 2016; Authors, 2015b; Van de Grift, 2014).

Activating teaching covers a learning environment in which students are aware of their learning, connecting it with their prior knowledge and using complex mental processes (Bonwell & Eison, 1991). Several studies have stated that this domain also affects relationships between students and with teachers (Authors, 2015b) and is theoretically and conceptually consistent with Kugel's (1993) 'learning' phase; it should be noted that this idea of teachers as guides or facilitators who monitor students' discussions requires additional significant and reflective tasks, which need a new conception of learners too (Gargallo et al., 2020).

An effective teacher should use varied teaching techniques in order to fit students' different learning styles. The use of metacognitive strategies has offered a framework in which students learn autonomously (also see the concept of 'agency' developed by Jensen et al., 2018: how the classroom milieu allows students to exercise choice, undertake responsibility, take on

different roles, and internalize learning expectations) and achieve more advanced learning skills (Authors, 2015b). Teachers and peers can also act as models providing other students with strategies to develop alternative ways to solve complex tasks (Kyriakides et al., 2013). Several indicators have reflected the use of these metacognitive processes: teaching how to simplify or order complex problems, asking learners to provide examples of their own, explaining how solutions can be applied in different situations or encouraging the use of alternative strategies. Student diversity requires adaptation to their individuality. According to Fuller's (1969) teachers' development model, this domain would be situated in the final and most complex stage (student concerns) which needs reflection about the impact of teaching on students' and teachers' ability to understand students' individual capacities.

Differentiation allows the inclusion of any student regardless of their ability (De Jager, 2011) and also the introduction of particular cultural dimensions - children's local knowledge and experiences outside of school- in classroom interactions (Jensen et al., 2018). Devoting extra time or resources, not focusing on the average learner, giving additional instructions to small groups of students and distinguishing between learners in terms of the length and size of assignments are just some examples of differentiated teaching. In all cases, correct and early diagnosis of students' academic problems or identification of at-risk students or minoritized students have been crucial (Jensen et al., 2018; Van de Grift, 2007).

It seems that there is a clear relationship between teacher teaching behaviours and student academic engagement. Furthermore, the more effective teaching behaviour exhibited by teachers, the better the student outcomes in terms of academic engagement (Skinner & Belmont, 1993; Woolley & Bowen, 2007) which can mediate their success, retention or motivation towards educational activities (Finn, 1993; Fredricks et al., 2004; Furrer & Skinner, 2003; Opdenakker & Minnaert, 2011; Skinner & Belmont, 1993). In short, students who are engaged in ongoing learning activities feel more pride and satisfaction in their

accomplishments and improve their competencies (Skinner & Belmont, 1993). Although the theoretical study of engagement has usually distinguished between behaviour, emotion and cognition, these factors have been dynamically interrelated within the individual and are not isolated processes (Fredricks et al., 2004).

Observation of teacher teaching behavior

Several procedures have been used to gather information about teacher practices. Even though other methods can be cheaper or more efficient, observation can give us a more accurate, objective, representative picture of the actual strategies adopted by teachers when these observational procedures fulfill certain standards (White, 2018). Moreover, according to Kelly et al. (2000) observational information has entailed growth in pedagogically relevant knowledge and can make teachers improve through experimenting with classroom practice and then reflecting on outcomes. Therefore, observation has now become more common (see Kelly et al., 2020 for a presentation of different observation protocols), but classroom dynamics are still interpreted by many teachers as a private space. Although accepting an external agent inside teachers' classroom has not always been easy, external observers are optimal as they can validly and reliably assess crucial features that students, for example, may not be qualified to judge (Hoyt & Pallet, 1999).

From a psychometric perspective, multiple and group observations would be best to avoid differences between observers, however organizational procedures inside schools have made this unrealistic (Van der Lans et al., 2016). Although evaluation outcomes based on one–time classroom observations have provided reliable insights about the specific lesson observed, it is also true that the presence of observers in the classrooms can influence teacher and student behaviour (De Jong & Westerhof, 2001).

The aim of this research has been to study teachers' teaching practices using an observation instrument, and to answer the following research questions:

What is the general standard of teachers' teaching behaviour as perceived by external observers?

What is the influence of the standard of teachers' teaching behaviour on student engagement?

Based on these questions and on previous empirical evidence regarding the continuous changes in Spanish educational legislation and the modest results in international studies, we have hypothesized:

H1. The perceived general standard of teacher behaviours will not be excellent, particularly in the most complex teaching domains.

H2. There will be differences in student engagement depending on the standard of teachers' teaching behavior.

Materials and methods

Sample

The participants have been 344 teachers in 56 public and private schools in the Principality of Asturias (Spain). All of them have been recruited based on voluntary participation in the study. Almost two-thirds of the teachers (214; 62.2%) have been women, 130 (37.8%) men. About a quarter (25.9%) taught languages, a quarter (25%) science and applied science, 18.3% social sciences, 17.4% vocational education and training subjects, 8.7% cultural and artistic education and 4.7% physical education. Teaching experience has ranged from less than five years (n = 34) to over 30 years (n = 42). 215 teachers (62.5%) gave their classes in lower secondary education, 66 (19.2%) in upper secondary education, and 63 (18.3%) in vocational education and training. Teachers worked in classes ranging in size from 2 to 35 students (M = 15.9; SD = 6.4). Interesting internal differences in class size have been found in the different educational stages: lower secondary education (M = 17.6; SD = 5.9); upper secondary education (M = 15.5; SD = 7); and vocational education and training (M = 10.9; SD = 4.9).

Instruments

The ICALT observation instrument, validated in several countries, has been used to gather information. The original instrument (based on Van de Grift, 2007 version) has been created in English so a back translation into Spanish following the procedure indicated by Hambleton et al. (2005) has been needed.

The instrument consists of 35 items grouped into seven domains: safe and stimulating learning climate (4 items), efficient classroom management (4 items), clarity of instruction (7 items), activating teaching (7 items), teaching - learning strategies (6 items), differentiation (4 items) and student academic engagement (3 items). The observers have recorded scores about effective teaching practices using a rating scale ranging from 1 (completely untrue) to 4 (completely true). For each high indicator, various examples of good practice have been given in order to establish a factual basis for the score. Scores of 1-2 indicate low quality teaching practices whereas scores of 3-4 represent high quality.

Procedure

To ensure that observations have been sufficiently objective and have followed the selected theoretical framework, only trained observers have carried out the fieldwork. Trainees have been organised in various sessions in order to avoid over-large groups. Secondary education teachers and the entire research team (54 observers) have participated in the observation training, which has involved broad information about the theoretical/methodological basis of the project and the observation instrument. In addition, the training has included an in–depth explanation of how to codify ratings using two 20-minute videos of secondary education lessons. Levels of observer agreement has been calculated and any confusing items have been discussed. Cut–off criteria for acceptable consensus among observers has been set at ≥ 0.70.

The observations have taken place in ordinary classes (excluding, for example, classes in which students were doing written tests) and have lasted for approximately 50-55 minutes.

Teacher behaviour has been recorded in real-time.

Data Analysis

Preliminary analyses to test the reliability and validity of the observation instrument have been performed. Firstly, a Confirmatory Factor Analysis (CFA) has been run in order to confirm the original structure (Van de Grift et al., 2014) using MPLUS 7.3 software. Maximum likelihood, has been selected as estimation method. The following measures have been included: the Chi-Square test of significance ($\chi 2$), the Tucker Lewis index–non normed fit index (TLI), the Comparative Fit Index (CFI), Steiger's Root Mean Square Error of Approximation (RMSEA) and the Standardized Root Mean Square Residual (SRMR).

In order to analyze the differential functioning of each of the items the paradigm of Item Response Theory has been employed using IRTPro Software 4 instead of the traditional Rash model. Consequently, values 1 -2 have been recoded to 0 (not perceived), and values 3-4 have been recoded to 1 (perceived) (Noben et al., 2020). The difficult level, the discrimination and fit indexes have also been run.

Cronbach's alpha has been used to assess the internal consistency of the scale and correlational analyses have also been performed to examine its relationship with student academic engagement as the criterion.

Descriptive analyses have been conducted to answer the first research question. To answer the second one, an analysis of variance has been performed. Students engagement dimension has been considered as criterion variable and teachers teaching behaviour standards predictor variables. In addition, the value of eta has been measured in order to determine the effect of teacher skills on student engagement. It should be noted that a $\eta 2$ value of about .01 has indicated a small effect size; between .06 and .10 a medium effect size; and over .14 a high effect size (Coe & Merino, 2003; Fritz & MacKinnon, 2007). When findings have revealed significant statistical differences within the criterion variable, post hoc analyses of difference

have also been carried out. A stepwise regression has been performed to identify which teaching skill domains have been revealed as more important when trying to promote student engagement.

Results

Psychometric quality of the observation instrument

Confirmatory factor analysis have exposed good fit of the model, $\chi^2 = 918.56$, df = 518, $p = .000$; TLI = 0.90, CFI = 0.91, RMSEA = .047 [.04, .05], SRMR = .07. Table 1 shows the descriptive statistics for each of the items. In our sample, Cronbach´s alpha has been .93. The reliability scores of the effective teaching behaviour domains have been: learning climate (.77), classroom management (.78), clarity of instruction (.81), activating teaching (.72), teaching–learning strategies (.86) differentiation (.72) and engagement (.88) indicating that all domains have been internally consistent.

Table 1.

The differential item functioning has shown a good fit: S-χ^2 higher than $p < .01$ (Edelen & Reeve, 2007) (Table 2). Data has revealed that those items related with learning climate and efficient classroom management have been more frequently observed whereas items which imply a higher level of difficulty (e.g. items inside differentiation domain) have been rarely observed. Focusing on those items which imply a lower level of difficulty, it can be said that item 7 “My teacher provides effective classroom management” and item 9 “My teacher presents and explains the subject material in a clear manner” have been the most frequently observed ones also obtaining the best discrimination indexes.

Table 2.

Table 3 illustrates the instrument validity in terms of mean inter-scale correlations, showing the correlation between the six domains and the criterion variable: student academic engagement.

Table 3

Positive, significant relationships between the six domains and student academic engagement have been found: learning climate ($r = 0.32$, $p < .01$), classroom management ($r = 0.46$, $p < .01$), clarity of instruction ($r = 0.48$, $p < .01$), activating teaching ($r = 0.51$, $p < .01$), teaching–learning strategies ($r = 0.45$, $p < .01$) and differentiation ($r = 0.31$, $p < .01$). The learning climate and differentiation domains have exhibited the weakest relationship with student academic engagement whereas activating teaching has shown the strongest relationship.

Mean inter-scale correlations have shown that, although a certain degree of overlap can be deduced by the values in some domains (e.g. activating teaching and clarity of instruction), the scales do seem to measure different teaching skills satisfactorily: learning climate (.41), classroom management (.44), clarity of instruction (.58), activating teaching (.60), teaching – learning strategies (.43) and differentiation (.37).

General level of observed teachers' teaching behaviour in Spain

Several categories have been established to classify the scores for effective teaching behaviour: unsatisfactory (1-2), satisfactory (2.01 - 3), good (3.01- 3.5) and excellent (3.51-4) (Authors, 2017a; Authors, 2015a). Scores have been converted to percentages to clarify differences in observations.

Table 4

In general, safe and stimulating learning climate, classroom management and clarity of instruction have been perceived as moderately strong whereas activating teaching, teaching learning strategies and differentiation moderately weak (Table 4). This pattern has suggested that on average, teachers have been seen by observers as better performers in the more basic effective teaching behaviours. It should be noted that the majority of teachers (40.7% and

41.6%) have demonstrated excellent performance in safe and stimulating learning climate and classroom management while the majority (45.1%) still lack mastery of differentiation behaviours, demonstrating unsatisfactory standards. On the other hand, 38.9%, 52.9% and 48.8% of the sample has exhibited satisfactory behaviour in clarity of instruction, activating teaching and teaching learning strategies. Hence, we can conclude that there is still room for improvement particularly in the less basic domains.

Influence of the standard of teachers' teaching behaviour on student engagement

An analysis of the parametric requirements has been performed in order to test the adequacy of the analysis of variance. All the variables have followed a normal distribution; skewness and kurtosis values have been lower than 1 in absolute values and the equality of variance has been verified by Levene's test. Because some criterion variables have not shown equality of variance in the predictor variables, Dunnet's C test has been selected for the post hoc analysis of variance.

In terms of engagement (Table 5), findings have shown significant statistical differences in the six domains with effect sizes between 22% and 8%. The highest effect sizes have been found in clarity of instruction and activating teaching (22%), followed by classroom management (19%) and teaching and learning strategies (18%). The lowest effect sizes have been found in safe and stimulating learning climate (9%) and differentiation (8%) which have shown medium values.

Table 5.

Post hoc analysis about safe and stimulating learning climate have exposed that differences have been focused on teachers with excellent skills and the remaining categories (Dunnet's C test is significant at .05). The same differences have been found for efficient classroom management. In the clarity of instruction domain, differences have been focused on teachers with unsatisfactory skills and the remaining categories, and between excellent teachers and

the rest. No differences have been observed between the middle categories (satisfactory and good). Two domains (activating teaching and teaching - learning strategies) have exhibited differences between all categories. Finally, in the differentiation domain, Dunnet's C test has shown differences between teachers with unsatisfactory skills and the remaining categories.

Once the effect sizes have been examined, the aim has been to identify which of the six teaching skill domains has been pointed out as the most important one for improving student engagement. To do so, a stepwise regression has been performed (Table 6). This information will make us able to identify which teaching learning domains can be key in future teacher training courses. Activating teaching, efficient classroom management and teaching-learning strategies have been the teaching skills which have demonstrated the highest predictive power (together they have explained 33% of the variance).

Table 6

Discussion and conclusions

This study has contributed to the validation of a model of teaching behaviour for the Spanish context as a way to improve learning opportunities inside classrooms and helping to provide a framework for instructional improvement grounded in knowledge and information about the teaching process (Kelly et al., 2020). According to Jensen et al. (2019) it has also clarified an explicit framework with clear unit(s) of analysis and theory based conjectures about how and why specific aspects of teaching affect student learning.

Our results have demonstrated that teachers have shown good levels of achievement in those basic skills (learning climate, classroom management and clarity of instruction) which seem to be a precondition for demonstrating outstanding levels in the other more complex domains (activating teaching, teaching learning strategies and differentiation). Our conclusions have also confirmed the cumulative order in the levels of difficulty of teachers'

teaching behaviours previously found by Authors (2014). In fact, the differential item functioning has revealed, in line with Noben et al., (2020), that items such us: “My teacher offers weaker learners extra study and instruction time” and “My teacher adjusts the processing of subject matter to relevant inter-learner differences” have been rarely observed and have shown the highest level of difficulty.

On the other hand, the observed teachers have not, on average, demonstrated excellent levels in the most complex domains, confirming Hypothesis 1. These findings have reinforced similar results found in previous phases of the research but using student ratings (Authors, 2019). Positive, significant relationships between all the domains and student academic engagement have been found, with learning climate and differentiation demonstrating the weakest relationships with student academic engagement, while activating teaching has revealed the strongest. Further research is needed about this fact, as only 42.2% of teachers have exhibited good/excellent standards in one of the domains which has shown a strong relationship with engagement (activating teaching). Consistently with previous studies (Bonwell & Eison, 1991; Christensen et al., 2009) this relationship is not surprising as this domain implies students’ awareness of their own learning, critical reflection after activities and the need to assume an active role as learners which necessarily means that the student is engaged and tries to cope with academic demands.

In the case of clarity of instruction, which also has a strong relationship with academic engagement, more than half the teachers (57.3%) have demonstrated good/excellent standards. Hence, as suggested by Cardwell (2011) teachers who are concerned about the quality of the education improve their students’ engagement. When the objectives of the lesson are clear, when teachers give good feedback, when they create learner assignments which stimulate active participation or make sure that all learners know what to do, observers have reported that students pay better attention and participate more actively.

Efficient classroom management and teaching learning strategies have also been important when the intention is to increase student engagement. Both of them mean that students have internalised a new way of learning transforming this concept into action: taking the initiative, working independently and assuming responsibility for their own learning process. As mentioned by Hospel and Galand (2016) this means a balance between teachers' autonomy support and structure. Obviously in this context and in line with Danielson (1996) and Oliver & Reschly (2007) the way in which teachers have organised classroom processes and their own strategies (e.g. emphasis on how to simplify complex problems, conscious encouragement to apply what has been learned, encouraging learners to think critically) have been revealed as decisive to improve students' engagement (Klem y Connell, 2004).

In contrast, as we have previously mentioned, domains such as a safe and stimulating learning climate and differentiation, although exhibiting a medium effect, have obtained between 7 and 14 points below the outstanding domains. Perhaps due to the low number of students with special needs in our sample (only 3.7% of the total number of students), the differentiation domain does not seem to have significantly determined student engagement in general. Further studies should be carried out to analyse the influence of this domain in this particular cohort of students. In the case of learning climate, as it is one of the more basic domains, one might conclude that its processes are too general to have a more noticeable influence on academic engagement.

Hypothesis 2 has also been confirmed. It is interesting to highlight that in the activating teaching and teaching - learning strategies domains, differences have been found between all categories of teaching skill standards. This fact can indicate that improvements in any of these teaching skills can increase student engagement, and their chances to improve achievement, reinforcing the conclusions obtained by Maulana & Opdenakker (2013) and Maulana et al. (2012, 2013). In contrast, differences in the effects of differentiation teaching skills have been

concentrated between unsatisfactory and the other three categories. To find differences in the highest levels other aspects need to be considered. To sum up, when a certain level has been reached in this domain, the analysis of potential improvements is more difficult. Our results have indicated that if teachers implement differentiation strategies at the right time, there will be an improvement in student engagement. On the other hand, higher standards in safe and stimulating learning climate and efficient classroom management are needed in order to improve student engagement. Finally, high standards in clarity of instruction will produce an effect on student engagement. In fact, when teachers do not address the skills associated with this domain, there will be negative effects.

To sum up, our results have suggested that the quality of teachers' teaching skills is an important predictor of students' academic engagement, which is in line with previous research already mentioned. We can conclude that the most determinant domains in order to improve student engagement have been: activating teaching so that students assume an active role; efficient classroom management which means among other things that there is supervision of how students perform their tasks and finally, teaching-learning strategies which means that students are taught how to simplify complex problems, how to test problem solutions, use check lists and think critically.

Support seems to be needed in order to help teachers attain sufficiently high standards. This in line with previous research (Danielson, 2012; Jensen et al., 2019; Reddy et al., 2013) which has shown that this support may be derived from an efficient feed-back from observation or student rating results, so that teachers would have the opportunity to reflect about their own practices and be benefited from information about specific effective instructional practices and how best to implement them. Observation can also help policy-makers improve educational practice by taking decisions based on information obtained from the observation of schools, providing a rational framework for improvement. This research can be helpful not

only for determining progress in classroom practice but also guiding initial and continual teacher training (Kelly et al., 2020). Special attention should be paid to the most complex behavioural domains which require long-term interventions in order to avoid large numbers of teachers never developing these more complex skills.

Limitations of the study

This study has several limitations. First, results have been based on one–time observations which can be substantially biased. Our results give us reliable insight into the specific lessons observed, but more observations would be needed to develop a more complete image of teachers' general teaching proficiency (Van der Lans et al., 2016). It would also be interesting to have multiple observers visiting the classrooms to give them some anonymity and protection in their ratings (Van der Lans et al., 2016). The current observation instrument should be mutually used together with alternative procedures for gathering student and teacher opinions, as long as using more than one source of information (triangulation) might mitigate the weakness of their individual perspectives.

Our sample may also be biased in that it has included teachers who have voluntarily decided to participate in the research. This means that teachers who perceive themselves as more effective, who feel comfortable in their classes or more motivated may have been overrepresented in the sample. Though it would be desirable in future research to increase the number of participants and randomly sample them.

References

Authors (2014).

Authors (2015a)

Authors (2015b)

Authors (2017a)

Authors (2017b)

Authors (2019)

Blaich, C., Wise, K., Pascarella, E. T., & Roksa, J. (2016). Instructional clarity and organization: it's not new or fancy, but it matters. *Change: The Magazine of Higher Learning*, *48*(4), 6-13. https://dx.doi.org/10.1080/00091383.2016.1198142

Bonwell, C. & Eison, J. A. (1991). *Active Learning: Creating Excitement in the Classroom. ASHE-ERIC Higher Education Reports No1*. The George Washington University, School of Education and Higher Education. Briole, S. & Maurin, E. (2019). Does evaluating teachers make a difference? *IZA Discussion Papers*, *123017*, 1 – 39. https://10.13140/RG.2.2.15662.33604

Briole, S. & Maurin, E. (2019). Does evaluating teachers make the difference? *Iza Discussion Paper Series, 12307,* 1 – 38.

Cardwell, M.E. (2011). Patterns of relationships between teacher engagement and student engagement. *Education Doctoral, 49,* 1 – 125.

Christensen, H., Vigild, M., Thomsen, E., Szabo, P., & Horsewell, A. (2009). *Activating teaching methods, studying responses and learning*. Arctic Technology Centre & Learning Lab DTU.

Coe, R., & Merino, C. (2003). Magnitud del efecto; Una guía para investigadores y usuarios. *Revista de Psicologıa de la PUCP, 21*, 146–177.

Danielson, C. (1996). *Enhancing Professional Practice: A Framework for Teaching.* Association for Supervision and Curriculum Development (ASCD).

Danielson, C. (2012). Observing classroom practice. *Educational Leadership, 70* (3), 32 – 37.

Danielson, C. (2013). Rubrics from the framework for teaching evaluation instrument. https://www.ode.state.or.us/wma/teachlearn/commoncore/danielson-2013-rubric-only.pdf

De Jager, T. (2011). Guidelines to assist the implementation of differentiated learning activities in South African secondary schools. *International Journal of Inclusive Education, 17*(1), 80-94. https://doi.org/10.1080/13603116.2011.580465

De Jong, R. & Westerhof, K.J. (2001). The quality of student ratings of teacher behaviour. *Learning Environments Research, 4*, 51–85. https://dx.doi.org/10.1023/A:1011402608575

Edelen M.O. & Reeve B.B. (2007). Applying item response theory (IRT) modeling to questionnaire development, evaluation, and refinement. *Quality of Life Research, 16* (1), 5-18. https// 10.1007/s11136-007-9198-0

Emmer, E. T. & Stough, L.M. (2001). Classroom management: a critical part of educational psychology, with implications for teacher education. *Educational Psychologist, 36*(2), 103-112. https://dx.doi.org/10.1207/S15326985EP3602_5

Finn, J. D. (1993). *School engagement and students at risk*. National Center for Educational Statistics.

Fredricks, J., Blumenfeld, P., & Paris, A. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research*, *74* (1), 59–109. https://dx.doi.org/ 10.3102/00346543074001059

Fritz, M., & MacKinnon, D. (2007). Required sample size to detect the mediated effect. *Psychological Science*, *12*, 233–238. https://dx.doi.org/10.1111/j.14678-9280.2007.01882.x

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. https://dx.doi.org/10.1037/0022-0663.95.1.148

Gage, N. A., Scott, T., Hirn, R. & MacSuga-Gage, A.S. (2018). The relationship between teachers' implementation of classroom management practices and student behavior in elementary school. *Behavioral Disorders, 43*(2), 302 – 315. https://doi.org/10.1177/0198742917714809

Gargallo, B., Pérez-Pérez, C., García-García, F. J., Giménez, J. A. y Portillo, N. (2020). La competencia aprender a aprender en la universidad: propuesta de modelo teórico. *Educación XX1, 23*(1), 19-44. https://doi.org/10.5944/educXX1.23367.

Hamre, B. K., Pianta, R. C., Downer, J. T., DeCoster, J., Mashburn, A. J., Jones, S. M., Brown, J. L., Cappella, E., Atkins, M., Rivers, S. E., Brackett, M. A., & Hamagami, A. (2013). Teaching through interactions: Testing a developmental framework of teacher effectiveness in over 4,000 classrooms. *The Elementary School Journal, 113*, 461–487. https://doi.org/10.1086/669616

Hambleton, R.K., Merenda, P., & Spielberger, C. (eds.) (2005). *Adapting educational and psychological tests for cross-cultural assessment*. Lawrence Erlbaum Publishers.

Hanushek, E. A. (2016). What matters for student achievement. *Education Next, 16*(2), 19-26.

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*. https://research.acer.edu.au/research_conference_2003/4/

Hospel, V. & Galand, B. (2016). Are both classroom autonomy support and structure equally important for students' engagement? A multilevel analysis. *Learning and Instruction, 41*, 1-10. https://doi.org/10.1016/j.learninstruc.2015.09.001.

Howe, C., Hennessy, S., Mercer, N., Vrikki, M., & Wheatley, L. (2019). Teacher–Student Dialogue During Classroom Teaching: Does It Really Impact on Student Outcomes?. *Journal of the Learning Sciences, 28* (4-5), 462-512. https://doi.org/10.1080/10508406.2019.1573730

Hoyt, D. P. & Pallet, W. H. (1999). Appraising teaching effectiveness: beyond student ratings. *Idea Paper, 36,* 1-14.

Jensen, B., Grajeda, S., & Haertel, E. (2018). Measuring cultural dimensions of classroom interactions. *Educational Assessment, 23*(4), 250-276.https://doi.org/10.1080/10627197.2018.1515010

Jensen, B., Wallace, T.L., Steinberg, M.P., Gabriel, R.E., Dietiker, L., Davis, D.S., Kelcey, B., Minor, E,, Halpin, P. & Rui, N. (2019). Complexity and scale in teaching effectiveness research: Reflections from the MET Study. *Education Policy Analysis Archives, 27*(7), 1 – 17. . http://dx.doi.org/10.14507/epaa.27.3923

Kelly, S., Bringe, R., Aucejo, E., & Fruehwirth, J. (2020). Using global observation protocols to inform research on teaching effectiveness and school improvement: Strengths and emerging limitations. *Education Policy Analysis Archives, 28*(62), 1 – 29..

Klem, A. M. & Connell, J. O. (2004). Relationships matter: linking teacher support to student engagement and achievement. *Journal of School Health, 74*(7), 262-273. https://doi.org/10.1111/j.1746-1561.2004.tb08283.x

Kugel, P. (1993). How professors develop as teachers. *Studies in Higher Education,* 18(3), 315 – 328.

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. https://dx.doi.org/10.1016/j.tate.2013.07.010

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. https://dx.doi.org/10.1016/j.tate.2008.06.001

Maulana, R., & Opdenakker, M.-C. (2013). Teachers' interpersonal involvement as a predictor of students' academic motivation among Indonesian secondary school students: A multilevel growth curve analysis. *The Asia-Pacific Education Researcher*, 23, 591-603. http://dx.doi.org/10.1007/s40299-013-0132-7

Maulana, R., Opdenakker, M.-C., & Bosker, R. (2013). Teacher–student interpersonal relationships do change and affect academic motivation: A multilevel growth curve modelling. *British Journal of Educational Psychology*. 84, 459-482.http://dx.doi.org/10.1111/bjep.12031

Maulana, R., Opdenakker, M.-C., Stroet, K., & Bosker, R. (2012). Observed lesson structure during the first year of secondary education: Exploration of changeand link with academic engagement. Teaching and Teacher Education, 28(6), 835–85. https://doi.org/10.1016/j.tate.2012.03.005

Muijs, D. & Reynolds, D. (2000). School effectiveness and teacher effectiveness in mathematics: some preliminary findings from the evaluation of the Mathematics Enhancement Programme (Primary*). School Effectiveness and School Improvement*, *11* (3), 273 – 303. https://dx.doi.org/10.1076/0924-3453(200009)11:3;1-G;FT273

Noben, I., Maulana, R., Deinum, J.F. & Adriaan-Hofman,W.H. (2020). Measuring university teachers' teaching quality: a Rasch modelling approach. *Learning Environments Research, 24,* 87 – 107. https://doi.org/10.1007/s10984-020-09319-w.

Oliver, R. M. & Reschly, D. J. (2007). *Effective classroom management: teacher preparation and professional development*. National Comprehensive center for Teacher Quality. https://files.eric.ed.gov/fulltext/ED543769.pdf

Opdenakker, M. V. & Minnaert, A. (2011). Relationships between learning environment characteristics and academic engagement. *Psychological Reports, 109*(1), 259-284. https://dx.doi.org/10.2466/09.10.11.PR0.109.4.259-284

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. https://dx.doi.org/10.3102/0013189X09332374

Reddy, L. A., & Dudek, C.M. (2014). Teacher progress monitoring of instructional and behavioral management practices: An evidence-based approach to improving classroom practices. *International Journal of School & Educational Psychology*, *2*, 71–84. https://doi.org/10.1080/21683603.2013.876951

Reddy, L.A., Fabiano, G., Dudek, C.M. & Hsu. L. (2013). Development and construct validity of the classroom strategies scale-observer form. *School Psychology Quarterly*, *28* (4), 317–341. https://doi.org/10.1037/spq0000043

Reyes, M.R., Brackett, M.A., Rivers, S.E., White, M. & Salovey, P. (2012). Classroom emotional climate, student engagement and academic achievement. *Journal of Educational Psychology*, 104, 700 -712. https://dx.doi.org/10.1037/a0027268

Sammons, P., & Bakkum, L. (2011). Effective schools, equity and teacher effectiveness: a review to the literature. *Profesorado, Revista de Curriculum y Formación del Profesorado, 15*(3), 9-26.

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. https://dx.doi.org/10.1037/0022-0663.85.4.571

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. https://dx.doi.org/ 10.1080/00131880701369651

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. https://dx.doi.org/10.1080/09243453.2013.794845

Van de Grift, W.; Helm – 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. http://dx.doi.org/10.1016/j.stueduc.2014.09.003

Van de Grift, W., & Lam, J. (1998). Het didactisch handelen in het basisonderwijs. Tijdschrift voor Onderwijsresearch, *23*(2), 224-241.

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. https://dx.doi.org/10.1111/emip.12078

Van der Lans, R. M., Van de Grift, W., Van Veen, K., & Fokkens-Bruinsma, M. (2016). Once is not enough: establishing reliability criteria for feedback and evaluation decision based on classroom observations. *Studies in Educational Evaluation, 50,* 88-95. https://dx.doi.org/10.1016/j.stueduc.2016.08.001

Wang, M. C., Haertel, G. D., & Walberg, H. J. (1993). Toward a Knowledge base for school learning. *Review of Educational Research*, *63*(3), 249-294. https://dx.doi.org/10.3102/00346543063003249

White, M. C. (2018). Rater performance standards for classroom observation measures. *Educational Researcher, 47*, 492–501. https://doi.org/10.3102/0013189X18785623

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. https://dx.doi.org/10.1111/j.1741-3729.2007.00442.x

Teaching behaviours in Spain under observation: an instrument for assessing teaching quality

La observación del comportamiento del profesorado: un instrumento para evaluar la calidad docente

7973 words

Acknowledgements

This work was supported by the Dutch scientific funding agency (NRO, under grant 405-15-732) and the Institute of Educational Research and Innovation of the University of Oviedo (INIE, under grant INIE-19- MOD C-1).

Abstract

Teaching quality affects student outcomes and is also reflected in the results obtained in international tests. Although several factors determine the results obtained by students in the already mentioned tests and their general academic performance, undoubtedly teachers constitute a key point. This study aims to describe teaching behaviours observed in 344 teachers in 56 public and private schools in Asturias (Spain). Descriptive analyses and analysis of variance have been run out in order to answer the research questions. A stepwise regression has been performed too, to identify which teaching skill domains are more important in promoting student engagement. Results have shown positive, significant relationships between all the domains and student academic engagement. Activating teaching, efficient classroom management and teaching and learning strategies seem to be the main teaching skills for increasing student engagement. Interesting differences have also been found in student engagement depending on the standard of teachers' teaching behavior.

Key words: observation; teaching quality; teaching effectiveness; assessment

Student learning depends on several factors which relate to students' own capabilities and motivation but also to their family background and teachers' skills. Undoubtedly, teachers are a key point in the learning process (Hanushek, 2016), and several theoretical models and authors (Danielson, 1996; De Jong & Westerhoff, 2001; Fuller, 1969; Hattie, 2003; Kyriakides et al., 2009; Kugel, 1993; Muijs & Reynolds, 2000; Pianta & Hamre, 2009; Sammons & Bakkun, 2011; Van de Grift, 2007) have explained the development of teachers' practices trying to identify which skills help teachers to be more effective and obtain better outcomes from their students.

Measuring teaching effectiveness

We have evidence (Briole & Maurin, 2019) that evaluating teachers generates significant benefits not only for teachers (e.g. improvement of their core skills) but also for students (e.g. improvement of educational equality).Theories of teacher development have been studied widely, resulting in different instruments for assessing teaching quality. Danielson (1996) has developed a model to assess effective teaching which has been divided into 22 components clustered in four domains of teaching responsibility: planning and preparation, classroom environment, instruction and professional responsibilities. In 2007 she developed a new framework which has resulted in a rubric from the Framework for Teaching Evaluation instrument (Danielson, 2013), including clear standards of practice.

Hamre and Pianta have also presented a scientific framework to assess classroom quality, the Classroom Assessment Scoring System (CLASS). This framework has assessed three domains of quality (emotional support, instructional support, and classroom organization) considering teacher–student interactions as being likely to contribute positively to students' development as a consequence of their experience in the classroom (Pianta & Hamre, 2009). In this same line, the teaching through interaction framework, has improved the CLASS system but considering the same three core domains (Hamre et al., 2013).

Also in the United States, Reddy & Dudek (2014), have stablished a system for teaching assessment including instructional and behavioral management strategies. The model, which has been based in the exhaustive revision of previous assessment models has considered whether strategies are used for individual students or groups, the inclusion of summaries of concepts, the presence of corrective feedback, the promotion of direct instruction, adaptive instruction or student thinking, to mention some examples (Reddy et al., 2013). This system has combined external observations of teachers and self-evaluation (Reddy et al., 2013) with the intention to promote a dialogue between teachers and observers.

In the European context (Cyprus), research during the 1990s can be seen as the starting point for a dynamic model which has tried to describe the complex nature of educational effectiveness. The model has considered multiple factors which have an influence at different levels: student, teacher/classroom, school and educational system. Focusing on teachers, the model has set out observable instructional behaviours grouped into eight factors: orientation, structuring, questioning, teaching modelling, application, time management, teacher role in making the classroom a learning environment and classroom assessment (Kyriakides et al., 2009).

The education inspectorates in several European regions have also undertaken comparative analyses of effective teaching using the observation instrument developed by Van de Grift and Lam (1998), which is a reliable, valid way to compare educational data, regardless of cultural differences between countries. The International Comparative Analysis of Learning and Teaching Project (ICALT) has integrated this instrument and the state of the art on teaching effectiveness in its procedure, also including non-European countries and secondary education. The adaptation of the initial domains has resulted in a final structure of six domains: safe and stimulating learning climate, classroom management, clarity of instruction, activating teaching, teaching and learning strategies and differentiation. There is a graduation

of levels of complexity in these domains. All of them have impact on students but the most complex ones are not easily acquired or deployed by most teachers (authors, 2014). In spite of this, teaching development should not be seen as a succession of rigid stages: Van der Lans et al. (2015) have demonstrated that the least complex skills in more complex domains may precede the development of the most complex skills in other less complex domains. The six domains are discussed in more detail below.

A safe and stimulating learning climate is one of the core domains because of its influence on student learning results and engagement (Reyes et al., 2012). Moreover, this influence of school climate have determined 20-40% of student achievement according to Van de Grift (2007). Although consensus about the exact characteristics of this kind of learning climate has not been reached, Wang et al., (1993) have reinforced the importance of this idea when they conclude that variables that affect students directly are the most determinant ones. Therefore, other domains such as those included in Danielson's (1996) model - school and policy level- despite needing to be considered, do not seem to be so determinant. This domain has a clear connection with teachers and students interactions, which have been identified as one of the most important aspects of teachers' job and so determinant to understand students' achievement, attitudes and motivation (Hamre et al., 2013; Howe et al., 2019).

Efficient classroom management allows teachers to achieve their objectives more easily, as it includes time management, orderly presentation of content, ensuring that the lesson begins and ends on time, appropriate balance of individual/group activities and effectively dealing with student misbehavior (Danielson, 1996; Oliver & Reschly, 2007; Van de Grift, 2007). Emmer & Stough (2001) have also demonstrated that efficient classroom management can reduce behaviour and discipline problems, leaving more time for educational purposes and providing more opportunities to learn. Nevertheless, other researchers have related classroom management practices with student engagement but not with disruptive behaviours (Gage et

al., 2018). On similar lines, research has suggested that effective teachers spend 15% less time on management and organization and 50% more time on instruction and interactive activities (Van de Grift, 2007), so effectively managing time leaves more time for direct educational purposes.

Clarity of instruction entails aspects of instructional quality such as giving staged instructions, making clear whether an answer is right or wrong and regularly checking if learners have understood what the lesson is about. Kyriakides et al. (2009) have referred to this teachers' intention to order and organize lesson information, as structure. Hence, students will not learn as much as they could, if instructions are unclear (Authors, 2017b). This domain has been connected with the idea of significant learning in which students' prior knowledge must be considered, giving them the opportunity to meaningfully learn in their lessons. Additionally, it requires explicit objectives, information about mistakes and the procedures required during each class (Blaich et al., 2016; Authors, 2015b; Van de Grift, 2014).

Activating teaching covers a learning environment in which students are aware of their learning, connecting it with their prior knowledge and using complex mental processes (Bonwell & Eison, 1991). Several studies have stated that this domain also affects relationships between students and with teachers (Authors, 2015b) and is theoretically and conceptually consistent with Kugel's (1993) 'learning' phase; it should be noted that this idea of teachers as guides or facilitators who monitor students' discussions requires additional significant and reflective tasks, which need a new conception of learners too (Gargallo et al., 2020).

An effective teacher should use varied teaching techniques in order to fit students' different learning styles. The use of metacognitive strategies has offered a framework in which students learn autonomously (also see the concept of 'agency' developed by Jensen et al., 2018: how the classroom milieu allows students to exercise choice, undertake responsibility, take on

different roles, and internalize learning expectations) and achieve more advanced learning skills (Authors, 2015b). Teachers and peers can also act as models providing other students with strategies to develop alternative ways to solve complex tasks (Kyriakides et al., 2013). Several indicators have reflected the use of these metacognitive processes: teaching how to simplify or order complex problems, asking learners to provide examples of their own, explaining how solutions can be applied in different situations or encouraging the use of alternative strategies. Student diversity requires adaptation to their individuality. According to Fuller's (1969) teachers' development model, this domain would be situated in the final and most complex stage (student concerns) which needs reflection about the impact of teaching on students' and teachers' ability to understand students' individual capacities.

Differentiation allows the inclusion of any student regardless of their ability (De Jager, 2011) and also the introduction of particular cultural dimensions - children's local knowledge and experiences outside of school- in classroom interactions (Jensen et al., 2018). Devoting extra time or resources, not focusing on the average learner, giving additional instructions to small groups of students and distinguishing between learners in terms of the length and size of assignments are just some examples of differentiated teaching. In all cases, correct and early diagnosis of students' academic problems or identification of at-risk students or minoritized students have been crucial (Jensen et al., 2018; Van de Grift, 2007).

It seems that there is a clear relationship between teacher teaching behaviours and student academic engagement. Furthermore, the more effective teaching behaviour exhibited by teachers, the better the student outcomes in terms of academic engagement (Skinner & Belmont, 1993; Woolley & Bowen, 2007) which can mediate their success, retention or motivation towards educational activities (Finn, 1993; Fredricks et al., 2004; Furrer & Skinner, 2003; Opdenakker & Minnaert, 2011; Skinner & Belmont, 1993). In short, students who are engaged in ongoing learning activities feel more pride and satisfaction in their

accomplishments and improve their competencies (Skinner & Belmont, 1993). Although the theoretical study of engagement has usually distinguished between behaviour, emotion and cognition, these factors have been dynamically interrelated within the individual and are not isolated processes (Fredricks et al., 2004).

Observation of teacher teaching behavior

Several procedures have been used to gather information about teacher practices. Even though other methods can be cheaper or more efficient, observation can give us a more accurate, objective, representative picture of the actual strategies adopted by teachers when these observational procedures fulfill certain standards (White, 2018). Moreover, according to Kelly et al. (2000) observational information has entailed growth in pedagogically relevant knowledge and can make teachers improve through experimenting with classroom practice and then reflecting on outcomes. Therefore, observation has now become more common (see Kelly et al., 2020 for a presentation of different observation protocols), but classroom dynamics are still interpreted by many teachers as a private space. Although accepting an external agent inside teachers' classroom has not always been easy, external observers are optimal as they can validly and reliably assess crucial features that students, for example, may not be qualified to judge (Hoyt & Pallet, 1999).

From a psychometric perspective, multiple and group observations would be best to avoid differences between observers, however organizational procedures inside schools have made this unrealistic (Van der Lans et al., 2016). Although evaluation outcomes based on one–time classroom observations have provided reliable insights about the specific lesson observed, it is also true that the presence of observers in the classrooms can influence teacher and student behaviour (De Jong & Westerhof, 2001).

The aim of this research has been to study teachers' teaching practices using an observation instrument, and to answer the following research questions:

What is the general standard of teachers' teaching behaviour as perceived by external observers?

What is the influence of the standard of teachers' teaching behaviour on student engagement?

Based on these questions and on previous empirical evidence regarding the continuous changes in Spanish educational legislation and the modest results in international studies, we have hypothesized:

H1. The perceived general standard of teacher behaviours will not be excellent, particularly in the most complex teaching domains.

H2. There will be differences in student engagement depending on the standard of teachers' teaching behavior.

Materials and methods

Sample

The participants have been 344 teachers in 56 public and private schools in the Principality of Asturias (Spain). All of them have been recruited based on voluntary participation in the study. Almost two-thirds of the teachers (214; 62.2%) have been women, 130 (37.8%) men. About a quarter (25.9%) taught languages, a quarter (25%) science and applied science, 18.3% social sciences, 17.4% vocational education and training subjects, 8.7% cultural and artistic education and 4.7% physical education. Teaching experience has ranged from less than five years (n = 34) to over 30 years (n = 42). 215 teachers (62.5%) gave their classes in lower secondary education, 66 (19.2%) in upper secondary education, and 63 (18.3%) in vocational education and training. Teachers worked in classes ranging in size from 2 to 35 students (M = 15.9; SD = 6.4). Interesting internal differences in class size have been found in the different educational stages: lower secondary education (M = 17.6; SD = 5.9); upper secondary education (M = 15.5; SD = 7); and vocational education and training (M = 10.9; SD = 4.9).

Instruments

The ICALT observation instrument, validated in several countries, has been used to gather information. The original instrument (based on Van de Grift, 2007 version) has been created in English so a back translation into Spanish following the procedure indicated by Hambleton et al. (2005) has been needed.

The instrument consists of 35 items grouped into seven domains: safe and stimulating learning climate (4 items), efficient classroom management (4 items), clarity of instruction (7 items), activating teaching (7 items), teaching - learning strategies (6 items), differentiation (4 items) and student academic engagement (3 items). The observers have recorded scores about effective teaching practices using a rating scale ranging from 1 (completely untrue) to 4 (completely true). For each high indicator, various examples of good practice have been given in order to establish a factual basis for the score. Scores of 1-2 indicate low quality teaching practices whereas scores of 3-4 represent high quality.

Procedure

To ensure that observations have been sufficiently objective and have followed the selected theoretical framework, only trained observers have carried out the fieldwork. Trainees have been organised in various sessions in order to avoid over-large groups. Secondary education teachers and the entire research team (54 observers) have participated in the observation training, which has involved broad information about the theoretical/methodological basis of the project and the observation instrument. In addition, the training has included an in–depth explanation of how to codify ratings using two 20-minute videos of secondary education lessons. Levels of observer agreement has been calculated and any confusing items have been discussed. Cut–off criteria for acceptable consensus among observers has been set at ≥ 0.70.

The observations have taken place in ordinary classes (excluding, for example, classes in which students were doing written tests) and have lasted for approximately 50-55 minutes.

Teacher behaviour has been recorded in real-time.

Data Analysis

Preliminary analyses to test the reliability and validity of the observation instrument have been performed. Firstly, a Confirmatory Factor Analysis (CFA) has been run in order to confirm the original structure (Van de Grift et al., 2014) using MPLUS 7.3 software. Maximum likelihood, has been selected as estimation method. The following measures have been included: the Chi-Square test of significance ($\chi 2$), the Tucker Lewis index–non normed fit index (TLI), the Comparative Fit Index (CFI), Steiger's Root Mean Square Error of Approximation (RMSEA) and the Standardized Root Mean Square Residual (SRMR).

In order to analyze the differential functioning of each of the items the paradigm of Item Response Theory has been employed using IRTPro Software 4 instead of the traditional Rash model. Consequently, values 1 -2 have been recoded to 0 (not perceived), and values 3-4 have been recoded to 1 (perceived) (Noben et al., 2020). The difficult level, the discrimination and fit indexes have also been run.

Cronbach's alpha has been used to assess the internal consistency of the scale and correlational analyses have also been performed to examine its relationship with student academic engagement as the criterion.

Descriptive analyses have been conducted to answer the first research question. To answer the second one, an analysis of variance has been performed. Students engagement dimension has been considered as criterion variable and teachers teaching behaviour standards predictor variables. In addition, the value of eta has been measured in order to determine the effect of teacher skills on student engagement. It should be noted that a $\eta 2$ value of about .01 has indicated a small effect size; between .06 and .10 a medium effect size; and over .14 a high effect size (Coe & Merino, 2003; Fritz & MacKinnon, 2007). When findings have revealed significant statistical differences within the criterion variable, post hoc analyses of difference

have also been carried out. A stepwise regression has been performed to identify which teaching skill domains have been revealed as more important when trying to promote student engagement.

Results

Psychometric quality of the observation instrument

Confirmatory factor analysis have exposed good fit of the model, $\chi^2 = 918.56$, df = 518, $p = .000$; TLI = 0.90, CFI = 0.91, RMSEA = .047 [.04, .05], SRMR = .07. Table 1 shows the descriptive statistics for each of the items. In our sample, Cronbach´s alpha has been .93. The reliability scores of the effective teaching behaviour domains have been: learning climate (.77), classroom management (.78), clarity of instruction (.81), activating teaching (.72), teaching–learning strategies (.86) differentiation (.72) and engagement (.88) indicating that all domains have been internally consistent.

Table 1.

The differential item functioning has shown a good fit: S-χ^2 higher than $p < .01$ (Edelen & Reeve, 2007) (Table 2). Data has revealed that those items related with learning climate and efficient classroom management have been more frequently observed whereas items which imply a higher level of difficulty (e.g. items inside differentiation domain) have been rarely observed. Focusing on those items which imply a lower level of difficulty, it can be said that item 7 “My teacher provides effective classroom management” and item 9 “My teacher presents and explains the subject material in a clear manner” have been the most frequently observed ones also obtaining the best discrimination indexes.

Table 2.

Table 3 illustrates the instrument validity in terms of mean inter-scale correlations, showing the correlation between the six domains and the criterion variable: student academic engagement.

Table 3

Positive, significant relationships between the six domains and student academic engagement have been found: learning climate ($r = 0.32$, $p < .01$), classroom management ($r = 0.46$, $p < .01$), clarity of instruction ($r = 0.48$, $p < .01$), activating teaching ($r = 0.51$, $p < .01$), teaching–learning strategies ($r = 0.45$, $p < .01$) and differentiation ($r = 0.31$, $p < .01$). The learning climate and differentiation domains have exhibited the weakest relationship with student academic engagement whereas activating teaching has shown the strongest relationship.

Mean inter-scale correlations have shown that, although a certain degree of overlap can be deduced by the values in some domains (e.g. activating teaching and clarity of instruction), the scales do seem to measure different teaching skills satisfactorily: learning climate (.41), classroom management (.44), clarity of instruction (.58), activating teaching (.60), teaching – learning strategies (.43) and differentiation (.37).

General level of observed teachers' teaching behaviour in Spain

Several categories have been established to classify the scores for effective teaching behaviour: unsatisfactory (1-2), satisfactory (2.01 - 3), good (3.01- 3.5) and excellent (3.51-4) (Authors, 2017a; Authors, 2015a). Scores have been converted to percentages to clarify differences in observations.

Table 4

In general, safe and stimulating learning climate, classroom management and clarity of instruction have been perceived as moderately strong whereas activating teaching, teaching learning strategies and differentiation moderately weak (Table 4). This pattern has suggested that on average, teachers have been seen by observers as better performers in the more basic

effective teaching behaviours. It should be noted that the majority of teachers (40.7% and 41.6%) have demonstrated excellent performance in safe and stimulating learning climate and classroom management while the majority (45.1%) still lack mastery of differentiation behaviours, demonstrating unsatisfactory standards. On the other hand, 38.9%, 52.9% and 48.8% of the sample has exhibited satisfactory behaviour in clarity of instruction, activating teaching and teaching learning strategies. Hence, we can conclude that there is still room for improvement particularly in the less basic domains.

Influence of the standard of teachers' teaching behaviour on student engagement

An analysis of the parametric requirements has been performed in order to test the adequacy of the analysis of variance. All the variables have followed a normal distribution; skewness and kurtosis values have been lower than 1 in absolute values and the equality of variance has been verified by Levene's test. Because some criterion variables have not shown equality of variance in the predictor variables, Dunnet's C test has been selected for the post hoc analysis of variance.

In terms of engagement (Table 5), findings have shown significant statistical differences in the six domains with effect sizes between 22% and 8%. The highest effect sizes have been found in clarity of instruction and activating teaching (22%), followed by classroom management (19%) and teaching and learning strategies (18%). The lowest effect sizes have been found in safe and stimulating learning climate (9%) and differentiation (8%) which have shown medium values.

Table 5.

Post hoc analysis about safe and stimulating learning climate have exposed that differences have been focused on teachers with excellent skills and the remaining categories (Dunnet's C test is significant at .05). The same differences have been found for efficient classroom

management. In the clarity of instruction domain, differences have been focused on teachers with unsatisfactory skills and the remaining categories, and between excellent teachers and the rest. No differences have been observed between the middle categories (satisfactory and good). Two domains (activating teaching and teaching - learning strategies) have exhibited differences between all categories. Finally, in the differentiation domain, Dunnet's C test has shown differences between teachers with unsatisfactory skills and the remaining categories.

Once the effect sizes have been examined, the aim has been to identify which of the six teaching skill domains has been pointed out as the most important one for improving student engagement. To do so, a stepwise regression has been performed (Table 6). This information will make us able to identify which teaching learning domains can be key in future teacher training courses. Activating teaching, efficient classroom management and teaching-learning strategies have been the teaching skills which have demonstrated the highest predictive power (together they have explained 33% of the variance).

Table 6

Discussion and conclusions

This study has contributed to the validation of a model of teaching behaviour for the Spanish context as a way to improve learning opportunities inside classrooms and helping to provide a framework for instructional improvement grounded in knowledge and information about the teaching process (Kelly et al., 2020). According to Jensen et al. (2019) it has also clarified an explicit framework with clear unit(s) of analysis and theory based conjectures about how and why specific aspects of teaching affect student learning.

Our results have demonstrated that teachers have shown good levels of achievement in those basic skills (learning climate, classroom management and clarity of instruction) which seem to be a precondition for demonstrating outstanding levels in the other more complex

domains (activating teaching, teaching learning strategies and differentiation). Our conclusions have also confirmed the cumulative order in the levels of difficulty of teachers' teaching behaviours previously found by Authors (2014). In fact, the differential item functioning has revealed, in line with Noben et al., (2020), that items such us: "My teacher offers weaker learners extra study and instruction time" and "My teacher adjusts the processing of subject matter to relevant inter-learner differences" have been rarely observed and have shown the highest level of difficulty.

On the other hand, the observed teachers have not, on average, demonstrated excellent levels in the most complex domains, confirming Hypothesis 1. These findings have reinforced similar results found in previous phases of the research but using student ratings (Authors, 2019). Positive, significant relationships between all the domains and student academic engagement have been found, with learning climate and differentiation demonstrating the weakest relationships with student academic engagement, while activating teaching has revealed the strongest. Further research is needed about this fact, as only 42.2% of teachers have exhibited good/excellent standards in one of the domains which has shown a strong relationship with engagement (activating teaching). Consistently with previous studies (Bonwell & Eison, 1991; Christensen et al., 2009) this relationship is not surprising as this domain implies students' awareness of their own learning, critical reflection after activities and the need to assume an active role as learners which necessarily means that the student is engaged and tries to cope with academic demands.

In the case of clarity of instruction, which also has a strong relationship with academic engagement, more than half the teachers (57.3%) have demonstrated good/excellent standards. Hence, as suggested by Cardwell (2011) teachers who are concerned about the quality of the education improve their students' engagement. When the objectives of the lesson are clear, when teachers give good feedback, when they create learner assignments

which stimulate active participation or make sure that all learners know what to do, observers have reported that students pay better attention and participate more actively.

Efficient classroom management and teaching learning strategies have also been important when the intention is to increase student engagement. Both of them mean that students have internalised a new way of learning transforming this concept into action: taking the initiative, working independently and assuming responsibility for their own learning process. As mentioned by Hospel and Galand (2016) this means a balance between teachers' autonomy support and structure. Obviously in this context and in line with Danielson (1996) and Oliver & Reschly (2007) the way in which teachers have organised classroom processes and their own strategies (e.g. emphasis on how to simplify complex problems, conscious encouragement to apply what has been learned, encouraging learners to think critically) have been revealed as decisive to improve students' engagement (Klem y Connell, 2004).

In contrast, as we have previously mentioned, domains such as a safe and stimulating learning climate and differentiation, although exhibiting a medium effect, have obtained between 7 and 14 points below the outstanding domains. Perhaps due to the low number of students with special needs in our sample (only 3.7% of the total number of students), the differentiation domain does not seem to have significantly determined student engagement in general. Further studies should be carried out to analyse the influence of this domain in this particular cohort of students. In the case of learning climate, as it is one of the more basic domains, one might conclude that its processes are too general to have a more noticeable influence on academic engagement.

Hypothesis 2 has also been confirmed. It is interesting to highlight that in the activating teaching and teaching - learning strategies domains, differences have been found between all categories of teaching skill standards. This fact can indicate that improvements in any of these teaching skills can increase student engagement, and their chances to improve achievement,

reinforcing the conclusions obtained by Maulana & Opdenakker (2013) and Maulana et al. (2012, 2013). In contrast, differences in the effects of differentiation teaching skills have been concentrated between unsatisfactory and the other three categories. To find differences in the highest levels other aspects need to be considered. To sum up, when a certain level has been reached in this domain, the analysis of potential improvements is more difficult. Our results have indicated that if teachers implement differentiation strategies at the right time, there will be an improvement in student engagement. On the other hand, higher standards in safe and stimulating learning climate and efficient classroom management are needed in order to improve student engagement. Finally, high standards in clarity of instruction will produce an effect on student engagement. In fact, when teachers do not address the skills associated with this domain, there will be negative effects.

To sum up, our results have suggested that the quality of teachers' teaching skills is an important predictor of students' academic engagement, which is in line with previous research already mentioned. We can conclude that the most determinant domains in order to improve student engagement have been: activating teaching so that students assume an active role; efficient classroom management which means among other things that there is supervision of how students perform their tasks and finally, teaching-learning strategies which means that students are taught how to simplify complex problems, how to test problem solutions, use check lists and think critically.

Support seems to be needed in order to help teachers attain sufficiently high standards. This in line with previous research (Danielson, 2012; Jensen et al., 2019; Reddy et al., 2013) which has shown that this support may be derived from an efficient feed-back from observation or student rating results, so that teachers would have the opportunity to reflect about their own practices and be benefited from information about specific effective instructional practices and how best to implement them. Observation can also help policy-makers improve

educational practice by taking decisions based on information obtained from the observation of schools, providing a rational framework for improvement. This research can be helpful not only for determining progress in classroom practice but also guiding initial and continual teacher training (Kelly et al., 2020). Special attention should be paid to the most complex behavioural domains which require long-term interventions in order to avoid large numbers of teachers never developing these more complex skills.

Limitations of the study

This study has several limitations. First, results have been based on one–time observations which can be substantially biased. Our results give us reliable insight into the specific lessons observed, but more observations would be needed to develop a more complete image of teachers' general teaching proficiency (Van der Lans et al., 2016). It would also be interesting to have multiple observers visiting the classrooms to give them some anonymity and protection in their ratings (Van der Lans et al., 2016). The current observation instrument should be mutually used together with alternative procedures for gathering student and teacher opinions, as long as using more than one source of information (triangulation) might mitigate the weakness of their individual perspectives.

Our sample may also be biased in that it has included teachers who have voluntarily decided to participate in the research. This means that teachers who perceive themselves as more effective, who feel comfortable in their classes or more motivated may have been overrepresented in the sample. Though it would be desirable in future research to increase the number of participants and randomly sample them.

References

Authors (2014).

Authors (2015a)

Authors (2015b)

Authors (2017a)

Authors (2017b)

Authors (2019)

Blaich, C., Wise, K., Pascarella, E. T., & Roksa, J. (2016). Instructional clarity and organization: it's not new or fancy, but it matters. *Change: The Magazine of Higher Learning*, *48*(4), 6-13. https://dx.doi.org/10.1080/00091383.2016.1198142

Bonwell, C. & Eison, J. A. (1991). *Active Learning: Creating Excitement in the Classroom. ASHE-ERIC Higher Education Reports No1*. The George Washington University, School of Education and Higher Education. Briole, S. & Maurin, E. (2019). Does evaluating teachers make a difference? *IZA Discussion Papers*, *123017*, 1 – 39. https://10.13140/RG.2.2.15662.33604

Briole, S. & Maurin, E. (2019). Does evaluating teachers make the difference? *Iza Discussion Paper Series, 12307*, 1 – 38.

Cardwell, M.E. (2011). Patterns of relationships between teacher engagement and student engagement. *Education Doctoral, 49*, 1 – 125.

Christensen, H., Vigild, M., Thomsen, E., Szabo, P., & Horsewell, A. (2009). *Activating teaching methods, studying responses and learning*. Arctic Technology Centre & Learning Lab DTU.

Coe, R., & Merino, C. (2003). Magnitud del efecto; Una guía para investigadores y usuarios. *Revista de Psicologıa de la PUCP*, *21*, 146–177.

Danielson, C. (1996). *Enhancing Professional Practice: A Framework for Teaching.* Association for Supervision and Curriculum Development (ASCD).

Danielson, C. (2012). Observing classroom practice. *Educational Leadership, 70* (3), 32 – 37.

Danielson, C. (2013). Rubrics from the framework for teaching evaluation instrument. https://www.ode.state.or.us/wma/teachlearn/commoncore/danielson-2013-rubric-only.pdf

De Jager, T. (2011). Guidelines to assist the implementation of differentiated learning activities in South African secondary schools. *International Journal of Inclusive Education, 17*(1), 80-94. https://doi.org/10.1080/13603116.2011.580465

De Jong, R. & Westerhof, K.J. (2001). The quality of student ratings of teacher behaviour. *Learning Environments Research, 4*, 51–85. https://dx.doi.org/10.1023/A:1011402608575

Edelen M.O. & Reeve B.B. (2007). Applying item response theory (IRT) modeling to questionnaire development, evaluation, and refinement. *Quality of Life Research, 16* (1), 5-18. https// 10.1007/s11136-007-9198-0

Emmer, E. T. & Stough, L.M. (2001). Classroom management: a critical part of educational psychology, with implications for teacher education. *Educational Psychologist, 36*(2), 103-112. https://dx.doi.org/10.1207/S15326985EP3602_5

Finn, J. D. (1993). *School engagement and students at risk*. National Center for Educational Statistics.

Fredricks, J., Blumenfeld, P., & Paris, A. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research*, *74* (1), 59–109. https://dx.doi.org/ 10.3102/00346543074001059

Fritz, M., & MacKinnon, D. (2007). Required sample size to detect the mediated effect. *Psychological Science*, *12*, 233–238. https://dx.doi.org/10.1111/j.14678-9280.2007.01882.x

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. https://dx.doi.org/10.1037/0022-0663.95.1.148

Gage, N. A., Scott, T., Hirn, R. & MacSuga-Gage, A.S. (2018). The relationship between teachers' implementation of classroom management practices and student behavior in elementary school. *Behavioral Disorders, 43*(2), 302 – 315. https://doi.org/10.1177/0198742917714809

Gargallo, B., Pérez-Pérez, C., García-García, F. J., Giménez, J. A. y Portillo, N. (2020). La competencia aprender a aprender en la universidad: propuesta de modelo teórico. *Educación XX1, 23*(1), 19-44. https://doi.org/10.5944/educXX1.23367.

Hamre, B. K., Pianta, R. C., Downer, J. T., DeCoster, J., Mashburn, A. J., Jones, S. M., Brown, J. L., Cappella, E., Atkins, M., Rivers, S. E., Brackett, M. A., & Hamagami, A. (2013). Teaching through interactions: Testing a developmental framework of teacher effectiveness in over 4,000 classrooms. *The Elementary School Journal, 113*, 461–487. https://doi.org/10.1086/669616

Hambleton, R.K., Merenda, P., & Spielberger, C. (eds.) (2005). *Adapting educational and psychological tests for cross-cultural assessment*. Lawrence Erlbaum Publishers.

Hanushek, E. A. (2016). What matters for student achievement. *Education Next, 16*(2), 19-26.

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*. https://research.acer.edu.au/research_conference_2003/4/

Hospel, V. & Galand, B. (2016). Are both classroom autonomy support and structure equally important for students' engagement? A multilevel analysis. *Learning and Instruction, 41*, 1-10. https://doi.org/10.1016/j.learninstruc.2015.09.001.

Howe, C., Hennessy, S., Mercer, N., Vrikki, M., & Wheatley, L. (2019). Teacher–Student Dialogue During Classroom Teaching: Does It Really Impact on Student Outcomes?. *Journal of the Learning Sciences, 28* (4-5), 462-512. https://doi.org/10.1080/10508406.2019.1573730

Hoyt, D. P. & Pallet, W. H. (1999). Appraising teaching effectiveness: beyond student ratings. *Idea Paper, 36,* 1-14.

Jensen, B., Grajeda, S., & Haertel, E. (2018). Measuring cultural dimensions of classroom interactions. *Educational Assessment, 23*(4), 250-276.https://doi.org/10.1080/10627197.2018.1515010

Jensen, B., Wallace, T.L., Steinberg, M.P., Gabriel, R.E., Dietiker, L., Davis, D.S., Kelcey, B., Minor, E,, Halpin, P. & Rui, N. (2019). Complexity and scale in teaching effectiveness research: Reflections from the MET Study. *Education Policy Analysis Archives, 27*(7), 1 – 17. . http://dx.doi.org/10.14507/epaa.27.3923

Kelly, S., Bringe, R., Aucejo, E., & Fruehwirth, J. (2020). Using global observation protocols to inform research on teaching effectiveness and school improvement: Strengths and emerging limitations. *Education Policy Analysis Archives, 28*(62), 1 – 29..

Klem, A. M. & Connell, J. O. (2004). Relationships matter: linking teacher support to student engagement and achievement. *Journal of School Health, 74*(7), 262-273. https://doi.org/10.1111/j.1746-1561.2004.tb08283.x

Kugel, P. (1993). How professors develop as teachers. *Studies in Higher Education,* 18(3), 315 – 328.

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. https://dx.doi.org/10.1016/j.tate.2013.07.010

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. https://dx.doi.org/10.1016/j.tate.2008.06.001

Maulana, R., & Opdenakker, M.-C. (2013). Teachers' interpersonal involvement as a predictor of students' academic motivation among Indonesian secondary school students: A multilevel growth curve analysis. *The Asia-Pacific Education Researcher,* 23, 591-603. http://dx.doi.org/10.1007/s40299-013-0132-7

Maulana, R., Opdenakker, M.-C., & Bosker, R. (2013). Teacher–student interpersonal relationships do change and affect academic motivation: A multilevel growth curve modelling. *British Journal of Educational Psychology*. 84, 459-482.http://dx.doi.org/10.1111/bjep.12031

Maulana, R., Opdenakker, M.-C., Stroet, K., & Bosker, R. (2012). Observed lesson structure during the first year of secondary education: Exploration of changeand link with academic engagement. Teaching and Teacher Education, 28(6), 835–85. https://doi.org/10.1016/j.tate.2012.03.005

Muijs, D. & Reynolds, D. (2000). School effectiveness and teacher effectiveness in mathematics: some preliminary findings from the evaluation of the Mathematics Enhancement Programme (Primary*). School Effectiveness and School Improvement*, *11* (3), 273 – 303. https://dx.doi.org/10.1076/0924-3453(200009)11:3;1-G;FT273

Noben, I., Maulana, R., Deinum, J.F. & Adriaan-Hofman,W.H. (2020). Measuring university teachers' teaching quality: a Rasch modelling approach. *Learning Environments Research, 24,* 87 – 107. https://doi.org/10.1007/s10984-020-09319-w.

Oliver, R. M. & Reschly, D. J. (2007). *Effective classroom management: teacher preparation and professional development*. National Comprehensive center for Teacher Quality. https://files.eric.ed.gov/fulltext/ED543769.pdf

Opdenakker, M. V. & Minnaert, A. (2011). Relationships between learning environment characteristics and academic engagement. *Psychological Reports, 109*(1), 259-284. https://dx.doi.org/10.2466/09.10.11.PR0.109.4.259-284

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. https://dx.doi.org/10.3102/0013189X09332374

Reddy, L. A., & Dudek, C.M. (2014). Teacher progress monitoring of instructional and behavioral management practices: An evidence-based approach to improving classroom practices. *International Journal of School & Educational Psychology, 2*, 71–84. https://doi.org/10.1080/21683603.2013.876951

Reddy, L.A., Fabiano, G., Dudek, C.M. & Hsu. L. (2013). Development and construct validity of the classroom strategies scale-observer form. *School Psychology Quarterly, 28* (4), 317–341. https://doi.org/10.1037/spq0000043

Reyes, M.R., Brackett, M.A., Rivers, S.E., White, M. & Salovey, P. (2012). Classroom emotional climate, student engagement and academic achievement. *Journal of Educational Psychology*, 104, 700 -712. https://dx.doi.org/10.1037/a0027268

Sammons, P., & Bakkum, L. (2011). Effective schools, equity and teacher effectiveness: a review to the literature. *Profesorado, Revista de Curriculum y Formación del Profesorado, 15*(3), 9-26.

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. https://dx.doi.org/10.1037/0022-0663.85.4.571

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. https://dx.doi.org/ 10.1080/00131880701369651

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. https://dx.doi.org/10.1080/09243453.2013.794845

Van de Grift, W.; Helm – 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. http://dx.doi.org/10.1016/j.stueduc.2014.09.003

Van de Grift, W., & Lam, J. (1998). Het didactisch handelen in het basisonderwijs. Tijdschrift voor Onderwijsresearch, *23*(2), 224-241.

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. https://dx.doi.org/10.1111/emip.12078

Van der Lans, R. M., Van de Grift, W., Van Veen, K., & Fokkens-Bruinsma, M. (2016). Once is not enough: establishing reliability criteria for feedback and evaluation decision based on classroom observations. *Studies in Educational Evaluation, 50*, 88-95. https://dx.doi.org/10.1016/j.stueduc.2016.08.001

Wang, M. C., Haertel, G. D., & Walberg, H. J. (1993). Toward a Knowledge base for school learning. *Review of Educational Research, 63*(3), 249-294. https://dx.doi.org/10.3102/00346543063003249

White, M. C. (2018). Rater performance standards for classroom observation measures. *Educational Researcher, 47*, 492–501. https://doi.org/10.3102/0013189X18785623

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. https://dx.doi.org/10.1111/j.1741-3729.2007.00442.x

Table 1.
Descriptive statistics in ICALT observation instrument

Domain	Item	Item Content	Mean	Variance	Loadings
		The teacher			
LC	1	… shows respect for learners in their behaviour and language.	3.71	0.32	.58
LC	2	… maintains a relaxed atmosphere.	3.55	0.47	.70
LC	3	… promotes learners' self-confidence.	3.20	0.76	.74
LC	4	… fosters mutual respect.	3.01	0.89	.66
EM	5	… ensures the lesson proceeds in an orderly manner.	3.34	0.61	.84
EM	6	… monitors to ensure learners carry out activities in the appropriate manner.	3.19	0.67	.54
EM	7	… provides effective classroom management.	3.47	0.52	.77
EM	8	… uses the time for learning efficiently .	3.46	0.55	.66
CI	9	… presents and explains the subject material in a clear manner.	3.36	0.57	.69
CI	10	… gives feedback to learners.	3.29	0.69	.57
CI	11	… engages all learners in the lesson.	3.13	0.64	.69
CI	12	… during the presentation stage, checks whether learners have understood the subject material.	3.18	0.81	.62
CI	13	… encourages learners to do their best.	2.54	1.05	.49
CI	14	… teaches in a well-structured manner.	3.23	0.67	.65
CI	15	… gives a clear explanation of how to use didactic aids and how to carry out assignments.	3.17	0.70	.60
AT	16	… offers activities and work forms that stimulate learners to take an active approach.	3.00	0.69	.49
AT	17	… stimulates the building of self-confidence in weaker leaners.	2.26	1.09	.45
AT	18	… stimulates learners to think about solutions.	2.76	0.66	.60
AT	19	… asks questions which stimulate learners to reflect.	3.05	0.63	.66
AT	20	… lets learners think aloud.	3.38	0.65	.50
AT	21	… gives interactive instructions.	3.28	0.61	.48
AT	22	… clearly specifies the lesson aims at the start of the lesson.	3.01	1.02	.51
DI	23	… evaluates whether the lesson aims have been reached.	2.46	1.15	.87
DI	24	… offers weaker learners extra study and instruction time.	1.90	0.98	.67
DI	25	… adjusts instructions to relevant inter-learner differences.	2.45	1.08	.67
DI	26	… adjusts the processing of subject matter to relevant inter-learner differences.	2.05	0.87	.78
TL	27	… teaches learners how to simplify complex problems.	2.64	1.18	.76
TL	28	… stimulates the use of control activities.	2.54	1.01	.77
TL	29	… teaches learners to check solutions.	2.63	0.89	.80
TL	30	… stimulates the application of what has been learned.	3.03	0.81	.61
TL	31	… encourages learners to think critically.	2.84	0.88	.55
TL	32	… asks learners to reflect on practical strategies.	2.42	1.01	.73
AE	33	… are fully engaged in the lesson.	3.30	0.60	.85
AE	34	… show that they are interested.	3.23	0.68	.93
AE	35	… take an active approach to learning	3.11	0.63	.75

Table 2.

Analysis of ICALT observation instrument

Dimension	Item	Descriptive Statistics					Fit Index	
		p_1	*b*	*SE*	*a*	*SE*	$S\text{-}\chi^2$	p_2
LC	1	.98	-2,78	0,56	1,78	0,58	3.51	.48
LC	2	.94	-2,38	0,42	1,44	0,37	16.41	.13
EM	8	.90	-1,87	0,26	1,63	0,34	13.34	.58
CI	10	.86	-1,77	0,27	1,29	0,26	21.22	.27
AT	20	.88	-1,72	0,23	1,56	0,31	18.19	.31
EM	7	.92	-1,62	0,17	2,55	0,53	16.23	.18
AT	21	.85	-1,61	0,22	1,48	0,29	17.28	.57
EM	5	.86	-1,6	0,21	1,65	0,32	10.35	.84
CI	9	.88	-1,52	0,17	2,16	0,41	15.26	.43
EM	6	.82	-1,48	0,21	1,34	0,25	18.10	.45
AE	33	.87	-1,41	0,15	2,25	0,42	18.66	.23
CI	14	.83	-1,37	0,17	1,65	0,3	13.52	.76
AE	35	.81	-1,26	0,16	1,62	0,29	14.87	.67
LC	3	.78	-1,23	0,18	1,38	0,24	19.26	.44
CI	11	.81	-1,2	0,14	1,89	0,32	19.12	.21
AE	34	.81	-1,18	0,14	1,95	0,34	15.83	.39
CI	15	.79	-1,17	0,15	1,68	0,29	21.68	.36
CI	12	.80	-1,15	0,14	1,84	0,31	15.59	.55
AT	16	.73	-1,12	0,19	1,12	0,2	20.13	.45
TL	30	,74	-1,1	0,17	1,21	0,22	19.26	.51
AT	19	.79	-1,05	0,12	1,99	0,33	17.17	.44
LC	4	.73	-0,99	0,16	1,3	0,22	21.23	.39
AT	22	.68	-0,86	0,17	1,08	0,19	17.99	.71
TL	31	.67	-0,68	0,12	1,48	0,24	22.92	.19
AT	18	.65	-0,67	0,13	1,28	0,21	15.65	.74
TL	27	.56	-0,25	0,1	1,79	0,27	13.04	.67
TL	29	.54	-0,19	0,11	1,44	0,22	24.11	.15
TL	28	.53	-0,13	0,11	1,37	0,21	19.07	.45
CI	13	.48	0,02	0,11	1,6	0,23	16.19	.51
DI	23	.48	0,02	0,14	0,99	0,17	21.46	.43
DI	25	.47	0,16	0,16	0,85	0,16	21.84	.41
TL	32	.45	0,16	0,11	1,68	0,25	10.85	.76
AT	17	.40	0,45	0,15	1,12	0,18	21.96	.23
DI	26	.27	1,13	0,21	1,07	0,19	16.67	.41
DI	24	.22	1,68	0,33	0,88	0,18	19.55	.24

Note. p_1 = Proportion of correct responses, *b* = Difficulty Level, *a* = Slope (Discrimination Index), $S\text{-}\chi^2$ = fit index, *SE* = Standard Error, p_2 = Significance level of $S\text{-}\chi^2$ statistic.

Table 3

Correlations between teachers' teaching behavior and student academic engagement

	LC	EM	CI	AT	TL	DI
Safe a stimulating learning climate (LC)						
Efficient classroom management (EM)	.39 **					
Clarity of instruction (CI)	.57 **	.66 **				
Activating teaching (AT)	.59 **	.52 **	.76 **			
Teaching-learning strategies (TL)	.29 **	.34 **	.51 **	.62 **		
Differentiation (DI)	.24**	.30 **	.44 **	.51 **	.38**	
Academic engagement (AE)	.32 **	.46 **	.48 **	.51 **	.45 **	.31 **
M	*M*	3.36	3.36	3.12	3.96	2.68
SD	0.60	0.59	0.58	0.53	0.75	0.75

* $p < .05$ ** $p < .01$ *** $p < .001$

Table 4

ICALT – domains scores

	Unsatisfactory *n* (%)	Satisfactory *n* (%)	Good *n* (%)	Excellent *n* (%)
Safe and stimulating learning climate	11(3.2%)	94(27.3%)	99 (28.8%)	140 (40.7)
Efficient classroom management	11 (3.2%)	104 (30.2%)	86 (25%)	143 (41.6%)
Clarity of instruction	15 (4.4%)	132(38.4%)	99 (28.8%)	98 (28.5%)
Activating teaching	17(4.9%)	182(52.9%)	88 (25.6%)	57(16.6%)
Teaching - learning strategies	78 (22.7%)	168(48.8%)	48(14%)	50(14.5%)
Differentiation	155 (45.1%)	147 (42.7%)	26(7. 6%)	15(4.4%)

Table 5.

Influence of the standard of teachers' teaching skills on student engagement

	Unsatisfactory		Satisfactory		Good		Excellent		F	η^2
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*		
Safe and stimulating learning climate	2.61	1.13	2.97	0.72	3.19	0.66	3.44	0.64	11.61***	.09
Efficient Classroom management	2.15	1.07	2.97	0.71	3.14	0.65	3.52	0.55	26.21***	.19
Clarity of instruction	2.07	1.02	3.04	0.65	3.25	0.64	3.59	0.53	31.22***	.22
Activating teaching	2.29	1.04	3.05	0.65	3.38	0.63	3.75	0.41	31.61***	.22
Teaching-learning strategies	2.81	0.86	3.16	0.60	3.44	0.62	3.79	0.44	25.54***	.18
Differentiation	3.02	0.71	3.31	0.71	3.58	0.54	3.69	0.48	9.73***	.08

Table 6

Predictors of student engagement

	Student engagement								
				Model 2			Model 3		
Variable	Model 1 *B*	95%IC	VIF	*B*	95%CI	VIF	*B*	95%CI	VIF
Constant	1.2	[0.83,1.57]		0.68	[0.27,1.04		0.67	[0.26,1.07]	
Activating teaching	0.68	[0.56,0.80]	1.00	0.50	[0.36,0.64]	1.38	0.34	[0.17,0.50]	1.98
Efficient classroom management				0.32	[0.19,0.44]	1.38	0.31	[0.19,0.43]	1.38
Teaching-learning strategies							0.19	[0.08,0.30]	1.62
R^2	.26			.31			.33		
F	119.27***			75.93***			56.35***		
ΔR^2	.26			.05			.02		
ΔF	119.27			24.43			12.20		

Note = CI = confidence interval. VIF = Variance Inflation Factor

* $p < .05$ ** $p < .01$ *** $p < .001$

Teaching behaviours in Spain under observation: an instrument for assessing teaching quality

La observación del comportamiento del profesorado: un instrumento para evaluar la calidad docente

7973 words

Acknowledgements

This work was supported by the Dutch scientific funding agency (NRO, under grant 405-15-732) and the Institute of Educational Research and Innovation of the University of Oviedo (INIE, under grant INIE-19- MOD C-1).

Abstract

~~Teaching quality affects student outcomes and is also reflected in the results obtained in international tests. Although several factors determine the results obtained by students in the already mentioned tests and their general academic performance, undoubtedly teachers constitute a key point~~. Teachers constitute a key point in the educational process. Therefore having information about their behaviours inside the classroom can give us very useful data when the intention is to improve teaching effectiveness. This study aims to describe teaching behaviours observed in 344 teachers in 56 public and private schools in Asturias (Spain). Descriptive analyses and analysis of variance have been run out in order to answer the research questions. A stepwise regression has been performed too, to identify which teaching skill domains are more important in promoting "student engagement". Results have shown positive, significant relationships between all the domains and "student academic engagement". "Activating teaching", "efficient classroom management" and "teaching-learning strategies" seem to be the main teaching skills for increasing "student engagement". Interesting differences have also been found in "student engagement" depending on the standard of teachers' teaching behaviour.

Key words: observation; teaching quality; teaching effectiveness; assessment

Students' learning, the results they accomplish and also their motivation depend on several factors which relate to students' own capabilities and classroom dynamics but also to their family background and teachers' skills. Undoubtedly, teachers are a key point in the learning process (Hanushek, 2016), and several theoretical models and authors (Danielson, 1996; De Jong & Westerhoff, 2001; Fuller, 1969; Hattie, 2003; Kyriakides et al., 2009; Kugel, 1993; Muijs & Reynolds, 2000; Pianta & Hamre, 2009; Sammons & Bakkun, 2011; Van de Grift, 2007) have explained the development of teachers' practices trying to identify which skills help teachers to be more effective and obtain better academic results ~~outcomes~~ from their students. Therefore, teachers' assessment seems to be crucial to obtain objective and feasible data concerning these issues.

Measuring teaching effectiveness

We have evidence (Briole & Maurin, 2019) that evaluating teachers may generate significant benefits not only for teachers (e.g. improvement of their core skills) but also for students (e.g. improvement of educational equality). When these positive effects are expected, evaluated teachers need a feedback so that they receive information about which of their behaviours inside the classroom need to be improved and which others, on the other hand, are outstanding. This is the unique way to link theoretical information about teachers' learning process and their practices, stimulating real opportunities to improve their daily tasks and competence (Tuytens & Devos, 2014). Although teachers' assessment is frequent is some levels of the Spanish educational system in other ones (e.g. secondary education) these kind of systematic procedures are scarce. The lack of this kind of systematic assessment about teaching behaviour in Spain, inspires one of the main motivations for this study.

Theories of teacher development have been studied widely, resulting in different instruments for assessing teachers' behaviours and teaching quality. Danielson (1996) has developed a model to assess effective teaching which has been divided into 22 components clustered in four

domains of teaching responsibility: planning and preparation, classroom environment, instruction and professional responsibilities. In 2007, she developed a new framework which has resulted in a rubric from the Framework for Teaching Evaluation instrument (Danielson, 2013) including clear standards of practice.

Pianta and Hamre have also presented a systematic ~~scientific~~ framework to assess classroom quality: the Classroom Assessment Scoring System (CLASS). This framework has assessed three domains of quality (emotional support, instructional support, and classroom organization) considering teacher–student interactions as being likely to contribute positively to students' development, as a consequence of their experience in the classroom (Pianta & Hamre, 2009). ~~In this same line,~~ The teaching through interaction framework has improved the CLASS system, but considering the same three core domains (Hamre et al., 2013).

Also in the United States, Reddy & Dudek (2014) have stablished a system for teaching assessment, including instructional and behavioural management strategies. The model, which has been based in the exhaustive revision of previous assessment models, has considered whether strategies are used for individual students or groups, the inclusion of summaries of concepts, the presence of corrective feedback, the promotion of direct instruction, adaptive instruction or student thinking, to mention some examples (Reddy et al., 2013). This system has combined external observations of teachers and self-evaluation (Reddy et al., 2013) with the intention to promote a dialogue between teachers and observers.

In the European context (Cyprus), research during the 1990s can be seen as the starting point for a dynamic model which has tried to describe the complex nature of educational effectiveness. The model has considered multiple factors which have an influence at different levels: student, teacher/classroom, school and educational system. Focusing on teachers the model has set out observable instructional behaviours, grouped into eight factors: orientation, structuring, questioning, teaching modelling, application, time management, teacher role in making the

classroom a learning environment and classroom assessment (Kyriakides et al., 2009).

The education inspectorates in several European regions have also undertaken comparative analyses of effective teaching, using the observation instrument developed by Van de Grift and Lam (1998). This instrument is a reliable, valid way to compare educational data, regardless of cultural differences between countries. The International Comparative Analysis of Learning and Teaching Project (ICALT) has integrated this instrument and the state of the art on teaching effectiveness in its procedure, also including non-European countries and secondary education. The adaptation of the initial domains has resulted in a final structure of six domains: "safe and stimulating learning climate", "efficient classroom management", "clarity of instruction", "activating teaching", "teaching-learning strategies" and "differentiation". It must be noted that there is a graduation of levels of complexity in these domains. All of them have impact on students, but the most complex ones are not easily acquired or deployed by most teachers (authors, 2014). In spite of this, teaching development should not be seen as a succession of rigid stages: Van der Lans et al. (2015) have demonstrated that the least complex skills in more complex domains may precede the development of the most complex skills in other less complex domains. The six domains will be discussed in more detail below.

Safe and stimulating learning climate

A "safe and stimulating learning climate" is one of the core domains, because of its influence on student learning results and engagement (Reyes et al., 2012). Moreover, this influence of school climate determines 20-40% of student achievement according to Van de Grift (2007). Although consensus about the exact characteristics of this kind of learning climate has not been reached, Wang et al. (1993) have reinforced the importance of this idea when they conclude that variables that affect students directly are the most determinant ones. Therefore, other domains such as those included in Danielson's (1996) model - school and policy level- despite needing to be considered, do not seem to be so determinant. This domain has a clear connection with the

interactions between teachers and students. These relationships have been identified as one of the most important aspects of teachers' job and seem so determinant to understand students' achievement, attitudes and motivation (Hamre et al., 2013; Howe et al., 2019).

Efficient classroom management

"Efficient classroom management" allows teachers to achieve their objectives more easily, as it includes time management, orderly presentation of content, ensuring that the lesson begins and ends on time, appropriate balance of individual/group activities and effectively dealing with student misbehaviour (Danielson, 1996; Oliver & Reschly, 2007; Van de Grift, 2007). Emmer & Stough (2001) have also demonstrated that "efficient classroom management" can reduce behaviour and discipline problems, leaving more time for educational purposes and providing more opportunities to learn. Nevertheless, other researchers have related classroom management practices with student engagement but not with disruptive behaviours (Gage et al., 2018). On similar lines, research has suggested that effective teachers spend 15% less time on management and organization, and 50% more time on instruction and interactive activities (Van de Grift, 2007). We can therefore conclude that effectively managing time, leaves more time for direct educational purposes.

Clarity of instruction

"Clarity of instruction" entails aspects of instructional quality such as giving staged instructions, making clear whether an answer is right or wrong, and regularly checking if learners have understood what the lesson is about. Hence, students will not learn as much as they could, if instructions are unclear (Authors, 2017b). Kyriakides et al. (2009) have referred to this teachers' intention to order and organize lesson information, as 'structure'. This domain has been connected with the idea of significant learning in which students' prior knowledge must be considered, giving them the opportunity to meaningfully learn in their lessons. Additionally, it requires explicit objectives, information about mistakes and, finally, the

procedures required during each class (Blaich et al., 2016; Authors, 2015b; Van de Grift, 2014).

Activating teaching

"Activating teaching" covers a learning environment in which students are aware of their learning connecting it with their prior knowledge, and using complex mental processes (Bonwell & Eison, 1991). Several studies have stated that this domain also affects relationships between students and with teachers (Authors, 2015b) and is theoretically and conceptually consistent with Kugel's (1993) 'learning' phase. It should be noted that this idea of teachers as guides, facilitators who monitor students' discussions, requires additional significant and reflective tasks which need a new conception of learners too (Gargallo et al., 2020).

Teaching-learning strategies

An effective teacher should use varied teaching techniques in order to fit students' different learning styles. The use of metacognitive strategies has offered a framework in which students learn autonomously (also see the concept of 'agency' developed by Jensen et al., 2018: how the classroom milieu allows students to exercise choice, undertake responsibility, take on different roles, and internalize learning expectations) and achieve more advanced learning skills (Authors, 2015b). Teachers and peers can also act as models, providing other students with strategies to develop alternative ways to solve complex tasks (Kyriakides et al., 2013). Several indicators can reflect the use of these metacognitive processes: teaching how to simplify or order complex problems, asking learners to provide examples of their own, explaining how solutions can be applied in different situations or encouraging the use of alternative strategies, to mention some examples. Moreover, student diversity also requires adaptation to their individuality and therefore the use of different teaching-learning strategies. According to Fuller's (1969) teachers' development model, this domain would be situated in the final and most complex stage (student concerns). This stage implies ~~which needs~~ reflection about the impact of teaching on students', and teachers' ability to understand students' individual capacities.

Differentiation

"Differentiation" allows the inclusion of any student regardless of his/her ability (De Jager, 2011) and also the introduction of particular cultural dimensions (children's local knowledge and experiences outside of school) in classroom interactions (Jensen et al., 2018). Devoting extra time or resources, not focusing on the average learner, giving additional instructions to small groups of students, and distinguishing between learners in terms of the length and size of assignments, are just some examples of differentiated teaching. In all cases, correct and early diagnosis of students' academic problems, ~~or~~ identification of at-risk ~~students~~ or minoritized students have been crucial (Jensen et al., 2018; Van de Grift, 2007).

It seems that there is a clear relationship between teacher teaching behaviours and "student academic engagement". Engagement refers to "the quality of a student's connection or involvement with the endeavour of schooling and hence with the people, activities, goals, values, and place that compose it" (Skinner, Kinderman et al., 2008, p. 494). The theoretical study of engagement has usually distinguished between behaviour, emotion and cognition (Connel & Wellborn, 1991; Skinner, Furrer, et al., 2008) as main domains to be considered inside engagement. Skinner, Kinderman et al. (2008) summarize all of them in two main domains: behavioural engagement and emotional one. These factors are dynamically interrelated within the individual, and are not isolated processes (Fredricks et al., 2004). Furthermore, the more effective teaching behaviour exhibited by teachers, the better the student outcomes in terms of "academic engagement" (Skinner & Belmont, 1993; Woolley & Bowen, 2007) which can mediate students' ~~their~~ success, retention or motivation towards educational activities (Finn, 1993; Fredricks et al., 2004; Furrer & Skinner, 2003; Opdenakker & Minnaert, 2011; Skinner & Belmont, 1993). Hence, this means that engagement seems to be a good predictor for students' future inside the educational system. In short, students who are engaged in ongoing learning activities feel more pride and satisfaction in their accomplishments, and

improve their competencies (Skinner & Belmont, 1993).

Observation of teacher teaching behavior

Several procedures have been used to gather information about teacher practices. Even though other methods can be cheaper or more efficient, observation can give us a more accurate, objective and representative picture of the actual strategies adopted by teachers when these observational procedures fulfil certain standards (White, 2018). Moreover, according to Kelly et al. (2000) observational information has entailed growth in pedagogically relevant knowledge, and can make teachers improve through experimenting with classroom practice and then reflecting on outcomes. Nevertheless, although observation has now become more common (see Kelly et al., 2020 for a presentation of different observation protocols), classroom dynamics are still interpreted by many teachers as a private space. Despite the fact that accepting an external agent inside teachers' classroom has not always been easy, external observers are optimal as they can validly and reliably assess crucial features that students, for example, may not be qualified to judge (Hoyt & Pallet, 1999).

From a psychometric perspective, multiple and group observations would be best to avoid differences between observers. However, organizational procedures inside schools have made this approach unrealistic (Van der Lans et al., 2016). Although evaluation outcomes based on one–time classroom observations can provide reliable insights about the specific lesson observed, it is also true that the presence of observers in the classrooms can influence teacher and student behaviour (De Jong & Westerhof, 2001).

The educational system in Spain

The Spanish government has passed various pieces of education legislation, trying to adapt the Spanish education system to today educational needs and to design a more flexible structure. In 1990, the Organic Law for the General Organization of the Educational System included, for the first time, secondary education as part of compulsory education in Spain. The main

consequence of this organizational change was that secondary education teachers started to be faced with more heterogeneous classes, motivations and different kinds of diverse ability levels than in the past, when only those students with successful results applied for it. Successive acts in 2002 (Organic Law of the Quality of Education), 2006 (Organic Law of Education), 2013 (Organic Law for the improvement of Quality of Education) and 2020 (Modification of Organic Law of Education) have changed the requirements for student assessment, the educational support received by students with special learning needs, and have tried to give the system a more european perspective (Consejo Escolar de Estado & Ministerio de Educación y Formación Profesional, 2018; García-Garrido, 2002; Puelles, 2008).

All of these legislative changes, the lack of agreement in education policy to maintain certain basic principles in the Spanish education system, together with changes in family socialization processes, have made education a contentious issue in Spain (Viñao, 2016). Perhaps as a consequence of this, Spanish results in international evaluations, such as the Programme for International Student Assessment (PISA), have not always been as good as hoped for, particularly in Mathematics (Ministerio de Educación, Cultura y Deporte de España, 2016; Organisation for Economic Cooperation and Development [OECD], 2018). Furthermore, there are substantial internal differences between Spanish regions with generally better results in all subjects in Castilla-León, Madrid, Navarra and Aragón (Ministerio de Educación, Cultura y Deporte de España, 2016). Although Spain has a central Department of Education that is responsible for coordination, educational responsibilities have been transferred to autonomous communities who are responsible for the design of curricula, language policies and other organizational issues concerning public schools, particularly in those regions which have their own language in addition to Spanish (Martínez-Usarralde, 2015).

The aim of this research has been to study teachers' teaching practices using an observation instrument and to answer the following research questions:

What is the general standard of teachers' teaching behaviour as perceived by external observers?

What is the influence of the standard of teachers' teaching behaviour on "student engagement"?

Based on these questions and on previous empirical evidence regarding the continuous changes in Spanish educational legislation and the modest results in international studies, we have hypothesized:

H1. The perceived general standard of teacher behaviours will be less outstanding in the most complex domains ("activating teaching", "teaching-leaning strategies" and "differentiation") than in simplest ones ("safe and stimulating learning climate", "efficient classroom management" and "clarity of instruction") ~~excellent (good practices teaching were showed upper 50%), particularly in the most complex teaching domains.~~

H2. There will be differences in student engagement depending on the standard of teachers' teaching behaviour.

Materials and methods

Sample

The participants were ~~have been~~ 344 teachers in 56 public and private schools in the Principality of Asturias (Spain). All of them were ~~have been~~ recruited based on voluntary participation in the study. Their involvement required prior authorization from the local education authorities and the head teachers of the schools. Teachers were given no remuneration for having being observed. These teachers taught in schools which had already participated in previous stages of the research (with questionnaires) or which had shown interest in joining the observation traineeship developed to assume the observation stage.

Almost two-thirds of the teachers (214; 62.2%) were ~~have been~~ women, 130 (37.8%) men. About a quarter (25.9%) taught languages, a quarter (25%) science and applied science, 18.3%

social sciences, 17.4% vocational education and training subjects, 8.7% cultural and artistic education and 4.7% physical education. Teaching experience ~~has~~ ranged from less than five years (n = 34) to over 30 years (n = 42). 215 teachers (62.5%) taught ~~gave their classes~~ in lower secondary education, 66 (19.2%) in higher secondary education, and 63 (18.3%) in vocational education and training. Teachers worked in classes ranging in size from 2 to 35 students (M = 15.9; SD = 6.4). Interesting internal differences in class size were ~~have been~~ found in the different educational stages: lower secondary education (M = 17.6; SD = 5.9); higher secondary education (M = 15.5; SD = 7); and vocational education and training (M = 10.9; SD = 4.9). All the students present in the class group of the volunteer observed teachers stayed during the observation process. There were few students with special educational needs in the observed classes (just 3.7% of the total number of students present in the classrooms during observations). Due to confidentiality matters, we did not record the classes or collect any individual information concerning students. Nevertheless, we did collect general information given by teachers about the general socioeconomic status (SES) and achievement of the groups, which were medium on average.

Instruments

The ICALT observation instrument, validated in several countries, was used to gather information. The original instrument (based on Van de Grift, 2007 version) had been created in English so a back translation into Spanish following the procedure indicated by Hambleton et al. (2005) was needed. With the intention to obtain a homogenous information in all the participants in the project. The Spanish research team selected the same instrument used in the other countries. Moreover, its previous use in so many countries gave more confidence in the success of the research purposes.

The instrument consisted of 35 items grouped into seven domains: “safe and stimulating learning climate” (4 items) (e.g. the teacher maintains a relaxed atmosphere), “efficient

classroom management" (4 items) (e.g. the teacher uses the time for learning efficiently),"clarity of instruction" (7 items) (e.g. the teacher teaches in well-structured manner), "activating teaching" (7 items) (e.g. the teacher stimulates the building of self-confidence in weaker leaners), "teaching - learning strategies" (6 items) (e.g. the teacher asks learners to reflect on practical strategies), "differentiation" (4 items) (e.g. the teacher offers weaker learners extra study and instruction time) and "student engagement" (3 items) (e.g. learners take an active approach to learning).

The observers recorded scores about effective teaching practices using a rating scale which ranged from 1 ("mostly weak") to 4 ("mostly strong") in the indicators. On the other hand, in the examples of good practice, answers varied from 0 ("no, I have not observed this") to 1 ("yes, I have observed this"). Therefore, for each high indicator, various examples of good practice (observed items) were given so that observers could establish a factual basis for the score (*e.g. Indicator: The teacher promotes learners' self-confidence. Examples of good practice: the teacher gives positive feedback on questions and remarks from leaners; the teacher compliments learners on their work; the teacher acknowledges the contributions that learners make*). Scores 1-2 indicated low quality teaching practices (indicators in which less than 50% of the examples of good practice were observed), whereas scores 3-4 represented high quality (indicators in which more than 50% of the examples of good practice were observed). The instrument also gave the option to include perceptions, worries, reflections or details observed during the session by the trained observers.

Subjects were categorized in six different groups: language; social science; science and applied sciences, physical education; cultural and artistic education and, finally, vocational education and training (VET) subjects.

Observed teachers' teaching experience considered 5 categories: less than 2 years of experience, 3-9 years of experience, 10-19 years of experience, 20-29 years of experience and

more than 30 years of experience.

Procedure

To ensure that observations were sufficiently objective and followed the selected theoretical framework, only trained observers carried out the fieldwork. The traineeship was developed in the University of Oviedo directed by some of the Dutch leaders of the ICALT project. Trainees were organised in various sessions in order to avoid over-large groups. Secondary education teachers and the entire research team (54 observers) participated in the observation training. The training involved broad information about the theoretical/methodological basis of the project and about the observation instrument. In addition, the training included an in–depth explanation of how to codify ratings using two 20-minute videos of secondary education lessons. Levels of observer agreement were calculated and any confusing items were discussed. Cut–off criteria for acceptable consensus among observers was set at ≥ 0.70.

47 trained teachers finally developed the observations. 37 of them were women (79%) and 10 men (21%). 4% of the observers had less than five years of teaching experience; 2% between 3-5 years of experience; 28% between 6-15 years of experience and, finally, 66% more than 15 years of teaching experience.

The observations took place in ordinary classes (excluding, for example, classes in which students were doing written tests) and lasted for approximately 50-55 minutes. Teachers' behaviour was recorded in real-time and the observation instrument was the same for all the disciplines.

Data Analysis

Preliminary analyses to test the reliability and validity of the observation instrument were performed. Firstly, a Confirmatory Factor Analysis (CFA) was run to confirm the original structure (Van de Grift et al., 2014) using MPLUS 7.3 software. Maximum likelihood, was selected as estimation method. The following measures were included: the Chi-Square test of

significance (χ2), the Tucker Lewis index–non normed fit index (TLI), the Comparative Fit Index (CFI), Steiger's Root Mean Square Error of Approximation (RMSEA) and the Standardized Root Mean Square Residual (SRMR).

In order to analyse the differential functioning of each of the items instead of the traditional Rash model the paradigm of Item Response Theory was employed, using IRTPro Software 4. Consequently, values 1-2 were recoded to 0 (not perceived) and values 3-4 to 1 (perceived) (Noben et al., 2020). The difficulty level, the discrimination and fit indexes were also run.

Cronbach's alpha was used to assess the internal consistency of the scale, and correlational analyses were also performed to examine its relationship with "student academic engagement" as the criterion.

Descriptive analyses were conducted to answer the first research question. To answer the second one, an analysis of variance was performed. The "student engagement" domain was considered as criterion variable and teachers teaching behaviour standards as predictor variables. In addition, the value of eta was measured in order to determine the effect of teacher skills on "student engagement". It should be noted that a η2 value of about .01 indicated a small effect size; between .06 and .10 a medium effect size; and over .14 a high effect size (Coe & Merino, 2003; Fritz & MacKinnon, 2007). When findings revealed significant statistical differences within the criterion variable, post hoc analyses of difference were also carried out. A stepwise regression was performed to identify which teaching skill domains were seen as more important when trying to promote "student engagement".

Results

Psychometric quality of the observation instrument

Confirmatory factor analysis exposed good fit of the model, $\chi^2 = 918.56$, df = 518, $p = .000$; TLI = 0.90, CFI = 0.91, RMSEA = .047 [.04, .05], SRMR = .07. Table 1 shows the descriptive statistics for each of the items. In our sample, Cronbach´s alpha was .93. The reliability scores

of the effective teaching behaviour domains were: "safe and stimulating learning climate" (.77), "efficient classroom management" (.78), "clarity of instruction" (.81), "activating teaching" (.72), "teaching–learning strategies" (.86), "differentiation" (.72) and "student engagement" (.88) indicating that all domains were internally consistent.

Table 1.

The differential item functioning showed a good fit: S-χ^2 higher than $p < .01$ (Edelen & Reeve, 2007) (Table 2). Data revealed that those items related with "safe and stimulating learning climate" and "efficient classroom management" were more frequently observed whereas items which imply a higher level of difficulty (e.g. items inside "differentiation" domain) were rarely observed. Focusing on those items which implied a lower level of difficulty, it can be said that item 7 "The teacher provides effective classroom management" and item 9 "The teacher presents and explains the subject material in a clear manner" were the most frequently observed ones. They also obtained the best discrimination indexes.

Table 2.

Table 3 illustrates the instrument validity in terms of mean inter-scale correlations, showing the correlation between the six domains and the criterion variable: "student academic engagement".

Table 3

Positive, significant relationships between the six domains and "student academic engagement" were found: "safe and stimulating learning climate" ($r = 0.32$, $p < .01$), "efficient classroom management" ($r = 0.46$, $p < .01$), "clarity of instruction" ($r = 0.48$, $p < .01$), "activating teaching" ($r = 0.51$, $p < .01$), "teaching–learning strategies" ($r = 0.45$, $p < .01$) and "differentiation" ($r = 0.31$, $p < .01$). The "safe and stimulating learning climate" and

"differentiation" domains exhibited the weakest relationship with "student engagement" whereas "activating teaching" showed the strongest one.

Mean inter-scale correlations showed that although a certain degree of overlap could be deduced by the values in some domains (e.g. "activating teaching" and "clarity of instruction"), the scales did seem to measure different teaching skills satisfactorily: "safe and stimulating learning climate" (.41), "efficient classroom management" (.44), "clarity of instruction" (.58), "activating teaching" (.60), "teaching-learning strategies" (.43) and "differentiation" (.37).

General level of observed teachers' teaching behaviour in Spain

Several categories were established to classify the scores for effective teaching behaviour: unsatisfactory (1-2), satisfactory (2.01 - 3), good (3.01- 3.5) and excellent (3.51- 4) (Authors, 2017a; Authors, 2015a). Scores were converted to percentages to clarify differences in observations.

Table 4

In general, "safe and stimulating learning climate", "efficient classroom management" and "clarity of instruction" were perceived as moderately strong. In the following notes we can clearly understand how do teachers develop these three domains:

- *Students are seated in a circle, with biscuits and drinks (tea and chocolate). All students represent a poem* [it is a group of 10 students] *they talk about the authors' life and even characterize themselves according to the historical context* (Female observer, more than 15 years of teaching experience, observing a female lower secondary education Language teacher)
- *The learning climate is quite good, there is peace, work and authority* (Female observer, between 6-15 years of teaching experience, observing a male higher secondary education English teacher)

- *Instructions are clearly understood by students* (Male observer, between 6-15 years of teaching experience, observing a lower secondary education female Geography teacher)
- *Students know and respect the rules without needing the teacher's intervention. The previous work developed during months to make these rues clear, can easily be perceived* (Female observer, more than 15 years of teaching experience, observing a male lower secondary education Physical Education teacher)
- *The tables and chair have been moved into a "U" shape to enhance collective learning during some activities* (Female observer, between 3-5 years of teaching experience, observing a female higher secondary education Biology teacher)

Nevertheless, one observer also pointed out the need to improve in one of these most basic domains, "safe and stimulating learning climate":

- *Students volume is too loud. They use a colloquial and inappropriate vocabulary, but the teacher takes no action* (Female observer, between 3-5 years of teaching experience, observing a female lower secondary education Chemistry teacher)

On the other hand, "activating teaching", "teaching-learning strategies" and "differentiation" were perceived as moderately weak (Table 4). This pattern suggested that, on average, teachers were seen by observers as better performers in the more basic effective teaching behaviours. Observers reflected about this fact in their "open" notes:

- *Three female students in the classroom have never been asked by the teacher during the classroom. None of them speaks or makes any questions, it seems that they are invisible for the teacher* (Female observer, more than 15 years of teaching experience, observing a male VET Technology teacher)
- *When daily problems are mentioned to stimulate reflection, some students find it difficult to connect the task with their social context. Some of them say that these examples are far from their real life* (Male observer, between 6-15 years of teaching experience, observing a male lower secondary education Geography teacher)

It should be noted that the majority of teachers (40.7% and 41.6%) demonstrated excellent performance in "safe and stimulating learning climate" and "classroom management" while the majority (45.1%) still lack mastery of "differentiation" behaviours, demonstrating unsatisfactory standards. Although as we have previously mentioned there were few students with special educational needs, observers pointed out teachers' behaviours concerning "efficient classroom management" and "safe and stimulating learning climate" which affected groups with this kind of students:

- *The group is formed by students with important learning difficulties (…) Two of them have been expelled from the school due to their bad behaviour. The teacher uses positive reinforcements very frequently* (Female observer, more than 15 years of teaching experience, observing a female lower secondary education English teacher)
- *It is a numerous group. There are important differences between students: there is a group with learning difficulties and another one which has adopted a very passive role. The firs ones, receive support of other teachers and the seconds do not have an easy solution. The diversity observed inside the classroom would require more support, additional teachers and other methodological measures* (Male observer, between 6-15 years of teaching experience, observing a male lower secondary education History teacher)

On the other hand, 38.9%, 52.9% and 48.8% of the sample exhibited satisfactory behaviour in "clarity of instruction", "activating teaching" and "teaching-learning strategies". These observers clearly perceived behaviours related to these domains during the session:

- *Students are organised, know what to do all the time* (Female observer, more than 15 years of teaching experience, observing a female higher secondary education Mathematics teacher)
- *It is a time scheduled lesson. The teacher uses an alarm* (Male observer, less than 3 years of teaching experience, observing a male lower secondary education Geography teacher)
- *The teacher uses TICs, gives a magisterial session and continuously asks students* (Female observer, more than 15 years of teaching experience, observing a female History teacher)

Hence, we can conclude that there is still room for improvement particularly in the less basic domains which imply more complex processes and a greater control of classroom dynamics. It also requires a profound knowledge about the students and the pedagogical requirements derived from their diversity.

Regarding subjects, differences were focused on "safe and stimulating learning climate" (F = 2.78, p = .02, η^2 = .04). Post hoc analysis revealed these differences between language subjects (M = 3.52, SD = 0.50) and VET subjects (M = 3.02, SD = 0.77) (d = -0. 496, p = .000); social sciences (M = 3.51, SD = 0.50) and VET subjects (d = -0. 479, p = .001). On the other hand, no statistical significant differences were found regarding teachers' teaching experience.

Finally, differences according school ownership were also analysed. Private school teachers developed more "activating teaching" strategies than those teachers from public ones (M = 3.1, SD = 0.52; vs M = 2.89, SD = 0.53; F = 5.78, p = .02, η^2 = .02). This finding was similar in "teaching-learning strategies" (private school teachers: M = 2.93, SD = 0.71; public school teacher: M = 2.60, SD = 0.71; F = 7.14, p = .01, η^2 = .02). Likewise, private school "students' engagement" was higher than public school students one (M = 3.40, SD = .67; vs M = 3.15, SD =.72; F = 4.63, p = .03, η^2 = .02).

Influence of the standard of teachers' teaching behaviour on student engagement

An analysis of the parametric requirements was performed in order to test the adequacy of the analysis of variance. All the variables followed a normal distribution; skewness and kurtosis values were lower than 1 in absolute values and the equality of variance was verified by Levene's test. Because some criterion variables did not show equality of variance in the predictor variables, Dunnet's C test was selected for the post hoc analysis of variance.

In terms of engagement (Table 5), findings showed significant statistical differences in the six domains with effect sizes between 22% and 8%. The highest effect sizes were found in "clarity of instruction" and "activating teaching" (22%), followed by "efficient classroom

management" (19%) and "teaching-learning strategies" (18%). The lowest effect sizes were found in "safe and stimulating learning climate" (9%) and "differentiation" (8%) which showed medium values. This observer gave her testimony about the difficulties found by some students to engage with the classroom:

- *Students show great difficulties to follow the lesson, they cannot concentrate. The lesson is very mechanic* (Female observer, more than 15 years of teaching experience, observing a female lower secondary education Music teacher)

Table 5.

Post hoc analysis about "safe and stimulating learning climate" exposed that differences were focused on teachers with excellent skills and the remaining categories (Dunnet's C test is significant at .05). The same differences were found for "efficient classroom management". In the "clarity of instruction" domain differences were focused on teachers with unsatisfactory skills and the remaining categories, and between excellent teachers and the rest. No differences were observed between the middle categories (satisfactory and good). Two domains ("activating teaching" and "teaching-learning strategies") exhibited differences between all categories. Finally, in the "differentiation" domain, Dunnet's C test showed differences between teachers with unsatisfactory skills and the remaining categories.

Once the effect sizes were examined, the aim was to identify which of the six teaching skill domains was pointed out as the most important one for improving "student engagement". To do so, a stepwise regression was performed (Table 6). This information would make us able to identify which teaching learning domains could be key in future teacher training courses. "Activating teaching", "efficient classroom management" and "teaching-learning strategies" were the teaching skills which demonstrated the highest predictive power (together they explained 33% of the variance).

Table 6

Discussion and conclusions

This study has contributed to the validation of a model of teaching behaviour for the Spanish context as a way to improve learning opportunities inside classrooms and helping to provide a framework for instructional improvement grounded in knowledge and information about the teaching process (Kelly et al., 2020). According to Jensen et al. (2019) it has also clarified an explicit framework with clear unit(s) of analysis and theory based conjectures, about how and why specific aspects of teaching affect student learning.

Our results demonstrated that teachers showed good levels of achievement in those basic skills ("safe and stimulating learning climate", "efficient classroom management" and "clarity of instruction") which seem to be a precondition for demonstrating outstanding levels in the other more complex domains ("activating teaching", "teaching-learning strategies" and "differentiation"). Nevertheless, observers have also found some exceptions showing inefficient behaviours, even in these basic domains. Our conclusions confirmed the cumulative order in the levels of difficulty of teachers' teaching behaviours previously found by Authors (2014). In fact, the differential item functioning revealed, in line with Noben et al. (2020), that items such us: "The teacher offers weaker learners extra study and instruction time" and "The teacher adjusts the processing of subject matter to relevant inter-learner differences" were rarely observed and showed the highest level of difficulty.

On the other hand, the observed teachers, on average, demonstrated less outstanding levels in the most complex domains, confirming Hypothesis 1. These findings reinforced similar results found in previous phases of the research but using student ratings (Authors, 2019). This idea has also been reflected in the observers' qualitative statements.

Positive, significant relationships between all the domains and "student academic engagement" were found. "Safe and stimulating learning climate" and "differentiation" were the domains which demonstrated the weakest relationships with "student academic

engagement". On the other hand, "activating teaching" revealed the strongest. Further research is needed about this fact, as only 42.2% of teachers exhibited good/excellent standards in one of the domains which showed a strong relationship with engagement ("activating teaching"). Consistently with previous studies (Bonwell & Eison, 1991; Christensen et al., 2009) this relationship is not surprising, as this domain implies students' awareness of their own learning, critical reflection after activities, and the need to assume an active role as learners which necessarily means that the student is engaged and tries to cope with academic demands.

In the case of "clarity of instruction", which also had a strong relationship with "academic engagement", more than half the teachers (57.3%) demonstrated good/excellent standards. Hence, as suggested by Cardwell (2011) teachers who were concerned about the quality of the education improved their "students' engagement". When the objectives of the lesson were clear, when teachers gave good feedback, when they created learner assignments which stimulate active participation or made sure that all learners knew what to do, observers reported that students paid better attention and participated more actively.

"Efficient classroom management" and "teaching-learning strategies" were also important when the intention was to increase student engagement. Both of them mean that students have internalised a new way of learning, transforming this concept into action: taking the initiative, working independently, and assuming responsibility for their own learning process. As mentioned by Hospel and Galand (2016) this means a balance between teachers' autonomy support and structure. In this context and in line with Danielson (1996) and Oliver & Reschly (2007), the way in which teachers organised classroom processes and their own strategies (e.g. emphasis on how to simplify complex problems, conscious encouragement to apply what has been learned, encouraging learners to think critically) was revealed as decisive to improve "students' engagement" (Klem & Connell, 2004).

In contrast, as we have previously mentioned, domains such as a "safe and stimulating learning climate" and "differentiation", although exhibiting a medium effect, obtained between 7 and 14 points below the outstanding domains. Perhaps due to the low number of students with special needs in our sample, the "differentiation" domain did not seem to have significantly determined student engagement in general. Further studies should be carried out to analyse the influence of this domain in this particular cohort of students. Even though the little presence of this kind of students is a fact, observers identified the attention they require as one of the main challenges teachers had to face with. In the case of "safe and stimulating learning climate", as it was one of the more basic domains, one might conclude that its processes were too general to have a more noticeable influence on "academic engagement".

Differences have been found according to subjects and the ownership of the school. These results reinforce the previously obtained conclusions using teachers and students questionnaires (Authors, 2021a). On the other hand, although in these previous stages differences were also found according teachers' teaching experience, we cannot confirm these data with these results derived from teachers' observation.

Hypothesis 2 has also been confirmed. It is interesting to highlight that in the "activating teaching" and "teaching-learning strategies" domains, differences were found between all categories of teaching skill standards. This fact can indicate that improvements in any of these teaching skills can increase "student engagement", and their chances to improve achievement, reinforcing the conclusions obtained by Maulana and Opdenakker (2013) and Maulana et al. (2012, 2013). In contrast, differences in the effects of "differentiation" teaching skills were concentrated between unsatisfactory and the other three categories. To find differences in the highest levels other aspects need to be considered. To sum up, when a certain level was reached in this domain, the analysis of potential improvements was more difficult. Our results indicated that if teachers implement "differentiation" strategies at the right time, there would be an

improvement in "student engagement". On the other hand, higher standards in "safe and stimulating learning climate" and "efficient classroom management" were needed in order to improve "student engagement". Finally, high standards in "clarity of instruction" will produce an effect on "student engagement". In fact, when teachers did not address the skills associated with this domain, there would be negative effects.

To sum up, our results have suggested that the quality of teachers' teaching skills was an important predictor of "students' academic engagement", which is in line with previous research already mentioned. We can conclude that the most determinant domains in order to improve "student engagement" were: "activating teaching" so that students assumed an active role; "efficient classroom management" which meant among other things that there was supervision of how students performed their tasks and finally, "teaching-learning" strategies which meant that students were taught how to simplify complex problems, how to test problem solutions, use check lists and think critically.

Support seems to be needed in order to help teachers attain sufficiently high standards. On the other hand, giving them a scientific and systematic information that proves that many teachers find it more difficult to achieve outstanding levels in some of these domains, may help them understand that it is not a "personal" problem but the complex consequence of their initial training or the resources stablished by the educational system to support their teaching procedures. This in line with previous research (Danielson, 2012; Jensen et al., 2019; Reddy et al., 2013) which showed that this support may be derived from an efficient feed-back from observation or student rating results, so that teachers would have the opportunity to reflect about their own practices and be benefited from information about specific effective instructional practices and how best to implement them.

Observation can also help policy-makers improve educational practice by taking decisions based on information obtained from the observation of schools and providing a rational

framework for improvement. As we have previously mentioned, this research can be helpful not only for determining progress in classroom practice, but also for guiding initial and continuous teacher training (Kelly et al., 2020). Special attention should be paid to the most complex behavioural domains which require long-term interventions, in order to avoid large numbers of teachers never developing these more complex skills. One of the main contributions offered both by this work and the ICALT project is the potentiality offered by the collective work developed in several countries. Hence, this homogenous approach can allow international comparisons (Authors, 2020; Authors 2021b) which may help in the improvement of worldwide teaching processes, while also considering national particularities.

Limitations of the study

This study has several limitations. First, results have been based on one–time observations which can be substantially biased. Our results gave us reliable insight into the specific lessons observed, but more observations would be needed to develop a more complete image of teachers' general teaching proficiency (Van der Lans et al., 2016). It would also be interesting to have multiple observers visiting the classrooms to give them some anonymity and protection in their ratings (Van der Lans et al., 2016).

The same observation instrument was used for different areas of knowledge and disciplines. Although the instrument focused on general domains which referred to planning, classroom climate or methodologies, tasks which may be observed in almost any teacher, in future research specific instruments should be designed to assure that particularities of all these subjects are sufficiently represented. One observer clearly pointed out the difficulties found when trying to use the instrument in some disciplines (Physical Education, in this case):

- *Items which have been given a rating of 1 or 2 refer to students' thinking or the time needed to understand. These items are difficult to apply in Physical Education, a subject in which action, reflexes and controlled time are determinant* (Female observer, more than 15 years

of teaching experience, observing a male lower secondary education Physical Education teacher)

The current observation instrument should be mutually used together with alternative procedures for gathering student and teacher opinions, as long as using more than one source of information (triangulation) might mitigate the weakness of their individual perspectives. Moreover, a more open instrument could also provide the opportunity to analyse and reflect about classroom processes, qualitative achievements, problem solutions or students' success. This alternative approach could give a more profound understanding of teaching-learning process and also promote more reflective analyses based on the strong and weak aspects which have been already detected.

Finally, our sample may also be biased in that it has included teachers who have voluntarily decided to participate in the research. This means that teachers who perceive themselves as more effective, who feel comfortable in their classes or more motivated, may have been overrepresented in the sample. Also due to this bias, in the observed classes (chosen by teachers between all the groups they taught) there were not many students with special educational needs, for example. Though, it would be desirable in future research to increase the number of participants, the selected groups and randomly sample them.

References

Authors (2014).

Authors (2015a)

Authors (2015b)

Authors (2017a)

Authors (2017b)

Authors (2019)

Authors (2020)

Authors (2021a)

Authors (2021b)

Blaich, C., Wise, K., Pascarella, E. T., & Roksa, J. (2016). Instructional clarity and organization: it's not new or fancy, but it matters. *Change: The Magazine of Higher Learning*, *48*(4), 6-13. https://dx.doi.org/10.1080/00091383.2016.1198142

Bonwell, C. & Eison, J. A. (1991). *Active Learning: Creating Excitement in the Classroom. ASHE-ERIC Higher Education Reports No1*. The George Washington University, School of Education and Higher Education. Briole, S. & Maurin, E. (2019). Does evaluating teachers make a difference? *IZA Discussion Papers*, *123017*, 1 – 39. https://10.13140/RG.2.2.15662.33604

Briole, S. & Maurin, E. (2019). Does evaluating teachers make the difference? *Iza Discussion Paper Series*, *12307*, 1 – 38.

Cardwell, M.E. (2011). Patterns of relationships between teacher engagement and student engagement. *Education Doctoral*, *49*, 1 – 125.

Christensen, H., Vigild, M., Thomsen, E., Szabo, P., & Horsewell, A. (2009). *Activating teaching methods, studying responses and learning*. Arctic Technology Centre & Learning Lab DTU.

Coe, R., & Merino, C. (2003). Magnitud del efecto; Una guía para investigadores y usuarios. *Revista de Psicologıa de la PUCP*, *21*, 146–177.

Connell, J. P., & Wellborn, J. G. (1991). Competence, autonomy, and relatedness: A motivational analysis of self-system processes. In M. R. Gunnar & L. A. Sroufe (Eds.), *Minnesota Symposia on Child Psychology. Self-processes and development* (pp. 43-77). Lawrence Erlbaum Associates.

Consejo Escolar de Estado & Ministerio de Educación y Formación Profesional (2018). *Informe 2018 sobre el estado del sistema educativo.* Ministerio de Educación y Formación

Profesional. http://www.educacionyfp.gob.es/educacion/mc/cee/publicaciones/informes-del-sistema-educativo/informe-2018.html

Danielson, C. (1996). *Enhancing Professional Practice: A Framework for Teaching.* Association for Supervision and Curriculum Development (ASCD).

Danielson, C. (2012). Observing classroom practice. *Educational Leadership, 70* (3), 32 – 37.

Danielson, C. (2013). Rubrics from the framework for teaching evaluation instrument. https://www.ode.state.or.us/wma/teachlearn/commoncore/danielson-2013-rubric-only.pdf

De Jager, T. (2011). Guidelines to assist the implementation of differentiated learning activities in South African secondary schools. *International Journal of Inclusive Education*, *17*(1), 80-94. https://doi.org/10.1080/13603116.2011.580465

De Jong, R. & Westerhof, K.J. (2001). The quality of student ratings of teacher behaviour. *Learning Environments Research*, *4*, 51–85. https://dx.doi.org/10.1023/A:1011402608575

Edelen M.O. & Reeve B.B. (2007). Applying item response theory (IRT) modeling to questionnaire development, evaluation, and refinement. *Quality of Life Research, 16* (1), 5-18. https// 10.1007/s11136-007-9198-0

Emmer, E. T. & Stough, L.M. (2001). Classroom management: a critical part of educational psychology, with implications for teacher education. *Educational Psychologist, 36*(2), 103-112. https://dx.doi.org/10.1207/S15326985EP3602_5

Finn, J. D. (1993). *School engagement and students at risk*. National Center for Educational Statistics.

Fredricks, J., Blumenfeld, P., & Paris, A. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research*, *74* (1), 59–109. https://dx.doi.org/10.3102/00346543074001059

Fritz, M., & MacKinnon, D. (2007). Required sample size to detect the mediated effect. *Psychological Science*, *12*, 233–238. https://dx.doi.org/10.1111/j.14678-9280.2007.01882.x

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. https://dx.doi.org/10.1037/0022-0663.95.1.148

Gage, N. A., Scott, T., Hirn, R. & MacSuga-Gage, A.S. (2018). The relationship between teachers' implementation of classroom management practices and student behavior in elementary school. *Behavioral Disorders, 43*(2), 302 – 315. https://doi.org/10.1177/0198742917714809

García-Garrido, J.L. (2002). La nueva Ley de Calidad: reflexiones en la perspectiva internacional. *Revista de Educación, 329,* 23 - 38.

Gargallo, B., Pérez-Pérez, C., García-García, F. J., Giménez, J. A. y Portillo, N. (2020). La competencia aprender a aprender en la universidad: propuesta de modelo teórico. *Educación XX1*, *23*(1), 19-44. https://doi.org/10.5944/educXX1.23367.

Hamre, B. K., Pianta, R. C., Downer, J. T., DeCoster, J., Mashburn, A. J., Jones, S. M., Brown, J. L., Cappella, E., Atkins, M., Rivers, S. E., Brackett, M. A., & Hamagami, A. (2013). Teaching through interactions: Testing a developmental framework of teacher effectiveness in over 4,000 classrooms. *The Elementary School Journal, 113,* 461–487. https://doi.org/10.1086/669616

Hambleton, R.K., Merenda, P., & Spielberger, C. (eds.) (2005). *Adapting educational and psychological tests for cross-cultural assessment*. Lawrence Erlbaum Publishers.

Hanushek, E. A. (2016). What matters for student achievement. *Education Next*, *16*(2), 19-26.

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*. https://research.acer.edu.au/research_conference_2003/4/

Hospel, V. & Galand, B. (2016). Are both classroom autonomy support and structure equally important for students' engagement? A multilevel analysis. *Learning and Instruction, 41*, 1-10. https://doi.org/10.1016/j.learninstruc.2015.09.001.

Howe, C., Hennessy, S., Mercer, N., Vrikki, M., & Wheatley, L. (2019). Teacher–Student Dialogue During Classroom Teaching: Does It Really Impact on Student Outcomes?. *Journal of the Learning Sciences, 28* (4-5), 462-512. https://doi.org/10.1080/10508406.2019.1573730

Hoyt, D. P. & Pallet, W. H. (1999). Appraising teaching effectiveness: beyond student ratings. *Idea Paper, 36*, 1-14.

Jensen, B., Grajeda, S., & Haertel, E. (2018). Measuring cultural dimensions of classroom interactions. *Educational Assessment, 23*(4), 250-276.https://doi.org/10.1080/10627197.2018.1515010

Jensen, B., Wallace, T.L., Steinberg, M.P., Gabriel, R.E., Dietiker, L., Davis, D.S., Kelcey, B., Minor, E,, Halpin, P. & Rui, N. (2019). Complexity and scale in teaching effectiveness research: Reflections from the MET Study. *Education Policy Analysis Archives, 27*(7), 1 – 17. . http://dx.doi.org/10.14507/epaa.27.3923

Kelly, S., Bringe, R., Aucejo, E., & Fruehwirth, J. (2020). Using global observation protocols to inform research on teaching effectiveness and school improvement: Strengths and emerging limitations. *Education Policy Analysis Archives, 28*(62), 1 – 29..

Klem, A. M. & Connell, J. O. (2004). Relationships matter: linking teacher support to student engagement and achievement. *Journal of School Health, 74*(7), 262-273. https://doi.org/10.1111/j.1746-1561.2004.tb08283.x

Kugel, P. (1993). How professors develop as teachers. *Studies in Higher Education,* 18(3), 315 – 328.

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. https://dx.doi.org/10.1016/j.tate.2013.07.010

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. https://dx.doi.org/10.1016/j.tate.2008.06.001

Martínez Usarralde, M.J. (2015). Spain. In W. Hörner, H. Döbert., L.R. Reuter, & B. Von Kopp (Eds.), *The Education Systems of Europe. Second edition* (pp. 759 – 777). Springer.

Maulana, R., & Opdenakker, M.-C. (2013). Teachers' interpersonal involvement as a predictor of students' academic motivation among Indonesian secondary school students: A multilevel growth curve analysis. *The Asia-Pacific Education Researcher*, 23, 591-603. http://dx.doi.org/10.1007/s40299-013-0132-7

Maulana, R., Opdenakker, M.-C., & Bosker, R. (2013). Teacher–student interpersonal relationships do change and affect academic motivation: A multilevel growth curve modelling. *British Journal of Educational Psychology*. 84, 459-482.http://dx.doi.org/10.1111/bjep.12031

Maulana, R., Opdenakker, M.-C., Stroet, K., & Bosker, R. (2012). Observed lesson structure during the first year of secondary education: Exploration of changeand link with academic engagement. Teaching and Teacher Education, 28(6), 835–85. https://doi.org/10.1016/j.tate.2012.03.005

Ministerio de Educación, Cultura y Deporte de España (2016). *PISA 2015. Programa para la evaluación internacional de los alumnos. Informe español.* Ministerio de Educacion, Cultura y Deporte.

Muijs, D. & Reynolds, D. (2000). School effectiveness and teacher effectiveness in mathematics: some preliminary findings from the evaluation of the Mathematics Enhancement Programme (Primary*). School Effectiveness and School Improvement, 11* (3), 273 – 303. https://dx.doi.org/10.1076/0924-3453(200009)11:3;1-G;FT273

Noben, I., Maulana, R., Deinum, J.F. & Adriaan-Hofman,W.H. (2020). Measuring university teachers' teaching quality: a Rasch modelling approach. *Learning Environments Research, 24*, 87 – 107. https://doi.org/10.1007/s10984-020-09319-w.

Oliver, R. M. & Reschly, D. J. (2007). *Effective classroom management: teacher preparation and professional development*. National Comprehensive center for Teacher Quality. https://files.eric.ed.gov/fulltext/ED543769.pdf

Opdenakker, M. V. & Minnaert, A. (2011). Relationships between learning environment characteristics and academic engagement. *Psychological Reports, 109*(1), 259-284. https://dx.doi.org/10.2466/09.10.11.PR0.109.4.259-284

Organisation for Economic Cooperation and Development (2018). *PISA 2015: PISA results in focus*. https://www.oecd.org/pisa/pisa-2015-results-in-focus.pdf

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. https://dx.doi.org/10.3102/0013189X09332374

Puelles, M. (2008). Las grandes leyes educativas de los últimos doscientos años. *Participación educativa, 7*, 7 - 15.

Reddy, L. A., & Dudek, C.M. (2014). Teacher progress monitoring of instructional and behavioral management practices: An evidence-based approach to improving classroom practices. *International Journal of School & Educational Psychology*, *2*, 71–84. https://doi.org/10.1080/21683603.2013.876951

Reddy, L.A., Fabiano, G., Dudek, C.M. & Hsu. L. (2013). Development and construct validity of the classroom strategies scale-observer form. *School Psychology Quarterly, 28* (4), 317–341. https://doi.org/10.1037/spq0000043

Reyes, M.R., Brackett, M.A., Rivers, S.E., White, M. & Salovey, P. (2012). Classroom emotional climate, student engagement and academic achievement. *Journal of Educational Psychology*, 104, 700 -712. https://dx.doi.org/10.1037/a0027268

Sammons, P., & Bakkum, L. (2011). Effective schools, equity and teacher effectiveness: a review to the literature. *Profesorado, Revista de Curriculum y Formación del Profesorado, 15*(3), 9-26.

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. https://dx.doi.org/10.1037/0022-0663.85.4.571

Skinner, E., Furrer, C., Marchand, G., & Kindermann, T. (2008). Engagement and disaffection in the classroom: Part of a larger motivational dynamic? *Journal of Educational Psychology, 100*(4), 765-781. https://10.1037/a0012840

Skinner, E.A., Kindermann, T.A., & Furrer, C.J. (2008). A motivational perspective on engagement and disaffection: Conceptualization and assessment of children's' behavioural and emotional participation in academic activities in the classroom. *Educational and Psychological Measurement, 69,* 493-525. http://10-1177/0013164408323233

Tuytens, M., & Devos, G. (2014). How to activate teachers through teacher evaluation. School *Effectiveness and School Improvement, 25*(4), 509-530, https://10.1080/09243453.2013.842601

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. https://dx.doi.org/ 10.1080/00131880701369651

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. https://dx.doi.org/10.1080/09243453.2013.794845

Van de Grift, W.; Helm – 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. http://dx.doi.org/10.1016/j.stueduc.2014.09.003

Van de Grift, W., & Lam, J. (1998). Het didactisch handelen in het basisonderwijs. Tijdschrift voor Onderwijsresearch, *23*(2), 224-241.

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. https://dx.doi.org/10.1111/emip.12078

Van der Lans, R. M., Van de Grift, W., Van Veen, K., & Fokkens-Bruinsma, M. (2016). Once is not enough: establishing reliability criteria for feedback and evaluation decision based on classroom observations. *Studies in Educational Evaluation, 50,* 88-95. https://dx.doi.org/10.1016/j.stueduc.2016.08.001

Viñao, A. (2016). La Ley Orgánica de Mejora de la Calidad Educativa (LOMCE) de 2013: ¿una reforma más? *Historia y Memoria de la Educación*, *3*, 137-170.

Wang, M. C., Haertel, G. D., & Walberg, H. J. (1993). Toward a Knowledge base for school learning. *Review of Educational Research*, *63*(3), 249-294. https://dx.doi.org/10.3102/00346543063003249

White, M. C. (2018). Rater performance standards for classroom observation measures. *Educational Researcher, 47*, 492–501. https://doi.org/10.3102/0013189X18785623

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. https://dx.doi.org/10.1111/j.1741-3729.2007.00442.x

Table 1.
Descriptive statistics in ICALT observation instrument

Domain	Item	Item Content	Mean	Variance	Loadings
		The teacher			
LC	1	… shows respect for learners in their behaviour and language.	3.71	0.32	.58
LC	2	… maintains a relaxed atmosphere.	3.55	0.47	.70
LC	3	… promotes learners' self-confidence.	3.20	0.76	.74
LC	4	… fosters mutual respect.	3.01	0.89	.66
EM	5	… ensures the lesson proceeds in an orderly manner.	3.34	0.61	.84
EM	6	… monitors to ensure learners carry out activities in the appropriate manner.	3.19	0.67	.54
EM	7	… provides effective classroom management.	3.47	0.52	.77
EM	8	… uses the time for learning efficiently .	3.46	0.55	.66
CI	9	… presents and explains the subject material in a clear manner.	3.36	0.57	.69
CI	10	… gives feedback to learners.	3.29	0.69	.57
CI	11	… engages all learners in the lesson.	3.13	0.64	.69
CI	12	… during the presentation stage, checks whether learners have understood the subject material.	3.18	0.81	.62
CI	13	… encourages learners to do their best.	2.54	1.05	.49
CI	14	… teaches in a well-structured manner.	3.23	0.67	.65
CI	15	… gives a clear explanation of how to use didactic aids and how to carry out assignments.	3.17	0.70	.60
AT	16	… offers activities and work forms that stimulate learners to take an active approach.	3.00	0.69	.49
AT	17	… stimulates the building of self-confidence in weaker leaners.	2.26	1.09	.45
AT	18	… stimulates learners to think about solutions.	2.76	0.66	.60
AT	19	… asks questions which stimulate learners to reflect.	3.05	0.63	.66
AT	20	… lets learners think aloud.	3.38	0.65	.50
AT	21	… gives interactive instructions.	3.28	0.61	.48
AT	22	… clearly specifies the lesson aims at the start of the lesson.	3.01	1.02	.51
DI	23	… evaluates whether the lesson aims have been reached.	2.46	1.15	.87
DI	24	… offers weaker learners extra study and instruction time.	1.90	0.98	.67
DI	25	… adjusts instructions to relevant inter-learner differences.	2.45	1.08	.67
DI	26	… adjusts the processing of subject matter to relevant inter-learner differences.	2.05	0.87	.78
TL	27	… teaches learners how to simplify complex problems.	2.64	1.18	.76
TL	28	… stimulates the use of control activities.	2.54	1.01	.77
TL	29	… teaches learners to check solutions.	2.63	0.89	.80
TL	30	… stimulates the application of what has been learned.	3.03	0.81	.61
TL	31	… encourages learners to think critically.	2.84	0.88	.55
TL	32	… asks learners to reflect on practical strategies.	2.42	1.01	.73
AE	33	… are fully engaged in the lesson.	3.30	0.60	.85
AE	34	… show that they are interested.	3.23	0.68	.93
AE	35	… take an active approach to learning	3.11	0.63	.75

Table 2.

Analysis of ICALT observation instrument

Dimension	Item	Descriptive Statistics					Fit Index	
		p_1	*b*	*SE*	*a*	*SE*	$S\text{-}\chi^2$	p_2
LC	1	.98	-2,78	0,56	1,78	0,58	3.51	.48
LC	2	.94	-2,38	0,42	1,44	0,37	16.41	.13
EM	8	.90	-1,87	0,26	1,63	0,34	13.34	.58
CI	10	.86	-1,77	0,27	1,29	0,26	21.22	.27
AT	20	.88	-1,72	0,23	1,56	0,31	18.19	.31
EM	7	.92	-1,62	0,17	2,55	0,53	16.23	.18
AT	21	.85	-1,61	0,22	1,48	0,29	17.28	.57
EM	5	.86	-1,6	0,21	1,65	0,32	10.35	.84
CI	9	.88	-1,52	0,17	2,16	0,41	15.26	.43
EM	6	.82	-1,48	0,21	1,34	0,25	18.10	.45
AE	33	.87	-1,41	0,15	2,25	0,42	18.66	.23
CI	14	.83	-1,37	0,17	1,65	0,3	13.52	.76
AE	35	.81	-1,26	0,16	1,62	0,29	14.87	.67
LC	3	.78	-1,23	0,18	1,38	0,24	19.26	.44
CI	11	.81	-1,2	0,14	1,89	0,32	19.12	.21
AE	34	.81	-1,18	0,14	1,95	0,34	15.83	.39
CI	15	.79	-1,17	0,15	1,68	0,29	21.68	.36
CI	12	.80	-1,15	0,14	1,84	0,31	15.59	.55
AT	16	.73	-1,12	0,19	1,12	0,2	20.13	.45
TL	30	,74	-1,1	0,17	1,21	0,22	19.26	.51
AT	19	.79	-1,05	0,12	1,99	0,33	17.17	.44
LC	4	.73	-0,99	0,16	1,3	0,22	21.23	.39
AT	22	.68	-0,86	0,17	1,08	0,19	17.99	.71
TL	31	.67	-0,68	0,12	1,48	0,24	22.92	.19
AT	18	.65	-0,67	0,13	1,28	0,21	15.65	.74
TL	27	.56	-0,25	0,1	1,79	0,27	13.04	.67
TL	29	.54	-0,19	0,11	1,44	0,22	24.11	.15
TL	28	.53	-0,13	0,11	1,37	0,21	19.07	.45
CI	13	.48	0,02	0,11	1,6	0,23	16.19	.51
DI	23	.48	0,02	0,14	0,99	0,17	21.46	.43
DI	25	.47	0,16	0,16	0,85	0,16	21.84	.41
TL	32	.45	0,16	0,11	1,68	0,25	10.85	.76
AT	17	.40	0,45	0,15	1,12	0,18	21.96	.23
DI	26	.27	1,13	0,21	1,07	0,19	16.67	.41
DI	24	.22	1,68	0,33	0,88	0,18	19.55	.24

Note. p_1 = Proportion of correct responses, *b* = Difficulty Level, *a* = Slope (Discrimination Index), $S\text{-}\chi^2$ = fit index, *SE* = Standard Error, p_2 = Significance level of $S\text{-}\chi^2$ statistic.

Table 3

Correlations between teachers' teaching behavior and student academic engagement

	LC	EM	CI	AT	TL	DI
Safe a stimulating learning climate (LC)						
Efficient classroom management (EM)	.39 **					
Clarity of instruction (CI)	.57 **	.66 **				
Activating teaching (AT)	.59 **	.52 **	.76 **			
Teaching-learning strategies (TL)	.29 **	.34 **	.51 **	.62 **		
Differentiation (DI)	.24**	.30 **	.44 **	.51 **	.38**	
Academic engagement (AE)	.32 **	.46 **	.48 **	.51 **	.45 **	.31 **
M	*M*	3.36	3.36	3.12	3.96	2.68
SD	0.60	0.59	0.58	0.53	0.75	0.75

* $p < .05$ ** $p < .01$ *** $p < .001$

Table 4

ICALT – domains scores

	Unsatisfactory *n* (%)	Satisfactory *n* (%)	Good *n* (%)	Excellent *n* (%)
Safe and stimulating learning climate	11(3.2%)	94(27.3%)	99 (28.8%)	140 (40.7)
Efficient classroom management	11 (3.2%)	104 (30.2%)	86 (25%)	143 (41.6%)
Clarity of instruction	15 (4.4%)	132(38.4%)	99 (28.8%)	98 (28.5%)
Activating teaching	17(4.9%)	182(52.9%)	88 (25.6%)	57(16.6%)
Teaching - learning strategies	78 (22.7%)	168(48.8%)	48(14%)	50(14.5%)
Differentiation	155 (45.1%)	147 (42.7%)	26(7. 6%)	15(4.4%)

Table 5.

Influence of the standard of teachers' teaching skills on student engagement

	Unsatisfactory		Satisfactory		Good		Excellent		F	η^2
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*		
Safe and stimulating learning climate	2.61	1.13	2.97	0.72	3.19	0.66	3.44	0.64	11.61***	.09
Efficient Classroom management	2.15	1.07	2.97	0.71	3.14	0.65	3.52	0.55	26.21***	.19
Clarity of instruction	2.07	1.02	3.04	0.65	3.25	0.64	3.59	0.53	31.22***	.22
Activating teaching	2.29	1.04	3.05	0.65	3.38	0.63	3.75	0.41	31.61***	.22
Teaching-learning strategies	2.81	0.86	3.16	0.60	3.44	0.62	3.79	0.44	25.54***	.18
Differentiation	3.02	0.71	3.31	0.71	3.58	0.54	3.69	0.48	9.73***	.08

Table 6

Predictors of student engagement

	Student engagement								
				Model 2			Model 3		
Variable	Model 1 *B*	95%IC	VIF	*B*	95%CI	VIF	*B*	95%CI	VIF
Constant	1.2	[0.83,1.57]		0.68	[0.27,1.04		0.67	[0.26,1.07]	
Activating teaching	0.68	[0.56,0.80]	1.00	0.50	[0.36,0.64]	1.38	0.34	[0.17,0.50]	1.98
Efficient classroom management				0.32	[0.19,0.44]	1.38	0.31	[0.19,0.43]	1.38
Teaching-learning strategies							0.19	[0.08,0.30]	1.62
R^2	.26			.31			.33		
F	119.27***			75.93***			56.35***		
ΔR^2	.26			.05			.02		
ΔF	119.27			24.43			12.20		

Note = CI = confidence interval. VIF = Variance Inflation Factor

* $p < .05$ ** $p < .01$ *** $p < .001$

Teaching behaviours in Spain under observation: an instrument for assessing teaching quality

La observación del comportamiento del profesorado: un instrumento para evaluar la calidad docente

7973 words

Acknowledgements

This work was supported by the Dutch scientific funding agency (NRO, under grant 405-15-732) and the Institute of Educational Research and Innovation of the University of Oviedo (INIE, under grant INIE-19- MOD C-1).

Abstract

Teachers constitute a key point in the educational process. Therefore having information about their behaviours inside the classroom can give us very useful data when the intention is to improve teaching effectiveness. This study aims to describe teaching behaviours observed in 344 teachers in 56 public and private schools in Asturias (Spain). Descriptive analyses and analysis of variance have been run out in order to answer the research questions. A stepwise regression has been performed too, to identify which teaching skill domains are more important in promoting "student engagement". Results have shown positive, significant relationships between all the domains and "student academic engagement". "Activating teaching", "efficient classroom management" and "teaching-learning strategies" seem to be the main teaching skills for increasing "student engagement". Interesting differences have also been found in "student engagement" depending on the standard of teachers' teaching behaviour.

Key words: observation; teaching quality; teaching effectiveness; assessment

Students' learning, the results they accomplish and also their motivation depend on several factors which relate to students' own capabilities and classroom dynamics but also to their family background and teachers' skills. Undoubtedly, teachers are a key point in the learning process (Hanushek, 2016) and according to the historical cultural approach teachers determine with their own behaviours that the selected teaching learning activities guide students' construction of learning in a certain direction (Coll,1988). Several theoretical models and authors (Danielson, 1996; De Jong & Westerhoff, 2001; Fuller, 1969; Hattie, 2003; Kyriakides et al., 2009; Kugel, 1993; Muijs & Reynolds, 2000; Pianta & Hamre, 2009; Sammons & Bakkun, 2011; Van de Grift, 2007) have explained the development of teachers' practices trying to identify which skills help teachers to be more effective and obtain better academic results from their students. Therefore, teachers' assessment seems to be crucial to obtain objective and feasible data concerning these issues. These processes can also stimulate in teachers' a reflection connecting their own knowledge about how to teach or treat their students with their previous experiences putting all this information 'in context'.

Measuring teaching effectiveness

We have evidence (Briole & Maurin, 2019) that evaluating teachers may generate significant benefits not only for teachers (e.g. improvement of their core skills) but also for students (e.g. improvement of educational equality). When these positive effects are expected, evaluated teachers need a feedback so that they receive information about which of their behaviours inside the classroom need to be improved and which others, on the other hand, are outstanding. This is one of the main ways to construct their own learning about teaching, assuming an active role, interpreting information and giving their experience a meaning (Cubero et al., 2007). Therefore it can connect theoretical information about teachers' learning process and their practices, stimulating real opportunities to improve their daily tasks and competence (Tuytens & Devos,

2014). Although teachers' assessment is frequent is some levels of the Spanish educational system in other ones (e.g. secondary education) these kind of systematic procedures are scarce. The lack of this kind of systematic assessment about teaching behaviour in Spain, inspires one of the main motivations for this study.

Theories of teacher development have been studied widely, resulting in different instruments for assessing teachers' behaviours and teaching quality. Danielson (1996) has developed a model to assess effective teaching which has been divided into 22 components clustered in four domains of teaching responsibility: planning and preparation, classroom environment, instruction and professional responsibilities. In 2007, she developed a new framework which has resulted in a rubric from the Framework for Teaching Evaluation instrument (Danielson, 2013) including clear standards of practice.

Pianta and Hamre have also presented a systematic framework to assess classroom quality: the Classroom Assessment Scoring System (CLASS). This framework has assessed three domains of quality (emotional support, instructional support, and classroom organization) considering teacher–student interactions as being likely to contribute positively to students' development, as a consequence of their experience in the classroom (Pianta & Hamre, 2009). The teaching through interaction framework has improved the CLASS system, but considering the same three core domains (Hamre et al., 2013).

Also in the United States, Reddy & Dudek (2014) have stablished a system for teaching assessment, including instructional and behavioural management strategies. The model, which has been based in the exhaustive revision of previous assessment models, has considered whether strategies are used for individual students or groups, the inclusion of summaries of concepts, the presence of corrective feedback, the promotion of direct instruction, adaptive instruction or student thinking, to mention some examples (Reddy et al., 2013). This system has combined external observations of teachers and self-evaluation (Reddy et al., 2013) with the

intention to promote a dialogue between teachers and observers.

In the European context (Cyprus), research during the 1990s can be seen as the starting point for a dynamic model which has tried to describe the complex nature of educational effectiveness. The model has considered multiple factors which have an influence at different levels: student, teacher/classroom, school and educational system. Focusing on teachers the model has set out observable instructional behaviours, grouped into eight factors: orientation, structuring, questioning, teaching modelling, application, time management, teacher role in making the classroom a learning environment and classroom assessment (Kyriakides et al., 2009).

The education inspectorates in several European regions have also undertaken comparative analyses of effective teaching, using the observation instrument developed by Van de Grift and Lam (1998). This instrument is a reliable, valid way to compare educational data, regardless of cultural differences between countries. The International Comparative Analysis of Learning and Teaching Project (ICALT) has integrated this instrument and the state of the art on teaching effectiveness in its procedure, also including non-European countries and secondary education. The adaptation of the initial domains has resulted in a final structure of six domains: "safe and stimulating learning climate", "efficient classroom management", "clarity of instruction", "activating teaching", "teaching-learning strategies" and "differentiation". It must be noted that there is a graduation of levels of complexity in these domains. All of them have impact on students, but the most complex ones are not easily acquired or deployed by most teachers (authors, 2014). In spite of this, teaching development should not be seen as a succession of rigid stages: Van der Lans et al. (2015) have demonstrated that the least complex skills in more complex domains may precede the development of the most complex skills in other less complex domains. The six domains will be discussed in more detail below.

Safe and stimulating learning climate

A "safe and stimulating learning climate" is one of the core domains, because of its influence

on student learning results and engagement (Reyes et al., 2012). Moreover, this influence of school climate determines 20-40% of student achievement according to Van de Grift (2007). Although consensus about the exact characteristics of this kind of learning climate has not been reached, Wang et al. (1993) have reinforced the importance of this idea when they conclude that variables that affect students directly are the most determinant ones. Therefore, other domains such as those included in Danielson's (1996) model - school and policy level- despite needing to be considered, do not seem to be so determinant. This domain has a clear connection with the interactions between teachers and students as long as classrooms constitute social defined groups and communities of practice (John-Steiner & Mahn, 1996) but also considers the educational importance attributed to relationships between students (Coll, 1984). These relationships have been identified as one of the most important aspects of teachers' job and seem so determinant to understand students' achievement, attitudes and motivation (Hamre et al., 2013; Howe et al., 2019).

Efficient classroom management

"Efficient classroom management" allows teachers to achieve their objectives more easily, as it includes time management, orderly presentation of content, ensuring that the lesson begins and ends on time, appropriate balance of individual/group activities and effectively dealing with student misbehaviour (Danielson, 1996; Oliver & Reschly, 2007; Van de Grift, 2007). Emmer & Stough (2001) have also demonstrated that "efficient classroom management" can reduce behaviour and discipline problems, leaving more time for educational purposes and providing more opportunities to learn. Nevertheless, other researchers have related classroom management practices with student engagement but not with disruptive behaviours (Gage et al., 2018). On similar lines, research has suggested that effective teachers spend 15% less time on management and organization, and 50% more time on instruction and interactive activities (Van de Grift, 2007). We can therefore conclude that effectively managing time, leaves more time for direct

educational purposes.

Clarity of instruction

"Clarity of instruction" entails aspects of instructional quality such as giving staged instructions, making clear whether an answer is right or wrong, and regularly checking if learners have understood what the lesson is about. Hence, students will not learn as much as they could, if instructions are unclear (Authors, 2017b). Kyriakides et al. (2009) have referred to this teachers' intention to order and organize lesson information, as 'structure'. This domain has been connected with the idea of significant learning in which students' prior knowledge must be considered, giving them the opportunity to meaningfully learn in their lessons starting with their zone of real development (Cole, 1984; Vygotsky, 1979). Additionally, it requires explicit objectives, information about mistakes and, finally, the procedures required during each class (Blaich et al., 2016; Authors, 2015b; Van de Grift, 2014).

Activating teaching

"Activating teaching" covers a learning environment in which students are aware of their learning connecting it with their prior knowledge, and using complex mental processes (Bonwell & Eison, 1991). Several studies have stated that this domain also affects relationships between students and with teachers (Authors, 2015b) and is theoretically and conceptually consistent with Kugel's (1993) 'learning' phase. According to Vygotsky (1979) this social interaction is determinant for learning. Therefore, learning is firstly a social issue which later develops into an individual one: the imitation of social activities guided by adults will allow students to reach more complex learning faster (Vygotsky, 1984). It should be noted that this idea of teachers as guides, facilitators who monitor students' discussions, requires additional significant and reflective tasks which need a new conception of learners too (Gargallo et al., 2020). Nevertheless, those teachers who support students learning more properly are those whose continuously adapt the type and grade of help to students' progress, promoting higher

levels of autonomy and control and who progressively assume a secondary role (Gispert & Onrubia, 1997).

Teaching-learning strategies

An effective teacher should use varied teaching techniques in order to fit students' different learning styles. The use of metacognitive strategies has offered a framework in which students learn autonomously (also see the concept of 'agency' developed by Jensen et al., 2018: how the classroom milieu allows students to exercise choice, undertake responsibility, take on different roles, and internalize learning expectations) and achieve more advanced learning skills (Carriedo, 1995; Authors, 2015b). Teachers and peers can also act as models, providing other students with strategies to develop alternative ways to solve complex tasks (Kyriakides et al., 2013). Several indicators can reflect the use of these metacognitive processes: teaching how to simplify or order complex problems, asking learners to provide examples of their own, explaining how solutions can be applied in different situations or encouraging the use of alternative strategies, to mention some examples. Moreover, student diversity also requires adaptation to their individuality and therefore the use of different teaching-learning strategies. According to Fuller's (1969) teachers' development model, this domain would be situated in the final and most complex stage (student concerns). This stage implies reflection about the impact of teaching on students', and teachers' ability to understand students' individual capacities. It also considers whether teachers have finally reached the meanings they have planned to share with students (Coll, 1988).

Differentiation

"Differentiation" allows the inclusion of any student regardless of his/her ability (De Jager, 2011) and also the introduction of particular cultural dimensions (children's local knowledge and experiences outside of school) in classroom interactions (Jensen et al., 2018). Devoting extra time or resources, not focusing on the average learner, giving additional instructions to

small groups of students, and distinguishing between learners in terms of the length and size of assignments, are just some examples of differentiated teaching. In all cases, correct and early diagnosis of students' academic problems, identification of at-risk or minoritized students have been crucial (Jensen et al., 2018; Van de Grift, 2007).

It seems that there is a clear relationship between teacher teaching behaviours and "student academic engagement". Engagement refers to "the quality of a student's connection or involvement with the endeavour of schooling and hence with the people, activities, goals, values, and place that compose it" (Skinner, Kinderman et al., 2008, p. 494). The theoretical study of engagement has usually distinguished between behaviour, emotion and cognition (Connel & Wellborn, 1991; Skinner, Furrer, et al., 2008) as main domains to be considered inside engagement. Skinner, Kinderman et al. (2008) summarize all of them in two main domains: behavioural engagement and emotional one. These factors are dynamically interrelated within the individual, and are not isolated processes (Fredricks et al., 2004). Furthermore, the more effective teaching behaviour exhibited by teachers, the better the student outcomes in terms of "academic engagement" (Skinner & Belmont, 1993; Woolley & Bowen, 2007) which can mediate students' ~~their~~ success, retention or motivation towards educational activities (Finn, 1993; Fredricks et al., 2004; Furrer & Skinner, 2003; Opdenakker & Minnaert, 2011; Skinner & Belmont, 1993). Hence, this means that engagement seems to be a good predictor for students' future inside the educational system. In short, students who are engaged in ongoing learning activities feel more pride and satisfaction in their accomplishments, and improve their competencies (Skinner & Belmont, 1993).

Observation of teacher teaching behaviour

Several procedures have been used to gather information about teacher practices. Even though other methods can be cheaper or more efficient, observation can give us a more accurate, objective and representative picture of the actual strategies adopted by teachers when these

observational procedures fulfil certain standards (White, 2018). Moreover, according to Kelly et al. (2000) observational information has entailed growth in pedagogically relevant knowledge, and can make teachers improve through experimenting with classroom practice and then reflecting on outcomes. Nevertheless, although observation has now become more common (see Kelly et al., 2020 for a presentation of different observation protocols), classroom dynamics are still interpreted by many teachers as a private space. Despite the fact that accepting an external agent inside teachers' classroom has not always been easy, external observers are optimal as they can validly and reliably assess crucial features that students, for example, may not be qualified to judge (Hoyt & Pallet, 1999).

From a psychometric perspective, multiple and group observations would be best to avoid differences between observers. However, organizational procedures inside schools have made this approach unrealistic (Van der Lans et al., 2016). Although evaluation outcomes based on one–time classroom observations can provide reliable insights about the specific lesson observed, it is also true that the presence of observers in the classrooms can influence teacher and student behaviour (De Jong & Westerhof, 2001).

The educational system in Spain

The Spanish government has passed various pieces of education legislation, trying to adapt the Spanish education system to today educational needs and to design a more flexible structure. In 1990, the Organic Law for the General Organization of the Educational System included, for the first time, secondary education as part of compulsory education in Spain. The main consequence of this organizational change was that secondary education teachers started to be faced with more heterogeneous classes, motivations and different kinds of diverse ability levels than in the past, when only those students with successful results applied for it. Successive acts in 2002 (Organic Law of the Quality of Education), 2006 (Organic Law of Education), 2013 (Organic Law for the improvement of Quality of Education) and 2020 (Modification of Organic

Law of Education) have changed the requirements for student assessment, the educational support received by students with special learning needs, and have tried to give the system a more European perspective (Consejo Escolar de Estado & Ministerio de Educación y Formación Profesional, 2018; García-Garrido, 2002; Puelles, 2008).

All of these legislative changes, the lack of agreement in education policy to maintain certain basic principles in the Spanish education system, together with changes in family socialization processes, have made education a contentious issue in Spain (Viñao, 2016). Perhaps as a consequence of this, Spanish results in international evaluations, such as the Programme for International Student Assessment (PISA), have not always been as good as hoped for, particularly in Mathematics (Ministerio de Educación, Cultura y Deporte de España, 2016; Organisation for Economic Cooperation and Development [OECD], 2018). Furthermore, there are substantial internal differences between Spanish regions with generally better results in all subjects in Castilla-León, Madrid, Navarra and Aragón (Ministerio de Educación, Cultura y Deporte de España, 2016). Although Spain has a central Department of Education that is responsible for coordination, educational responsibilities have been transferred to autonomous communities who are responsible for the design of curricula, language policies and other organizational issues concerning public schools, particularly in those regions which have their own language in addition to Spanish (Martínez-Usarralde, 2015).

The aim of this research has been to study teachers' teaching practices using an observation instrument and to answer the following research questions:

What is the general standard of teachers' teaching behaviour as perceived by external observers?

What is the influence of the standard of teachers' teaching behaviour on "student engagement"?

Based on these questions and on previous empirical evidence regarding the continuous

changes in Spanish educational legislation and the modest results in international studies, we have hypothesized:

H1. The perceived general standard of teacher behaviours will be less outstanding in the most complex domains ("activating teaching", "teaching-leaning strategies" and "differentiation") than in simplest ones ("safe and stimulating learning climate", "efficient classroom management" and "clarity of instruction")

H2. There will be differences in student engagement depending on the standard of teachers' teaching behaviour.

Materials and methods

Sample

The participants were 344 teachers in 56 public and private schools in the Principality of Asturias (Spain). All of them were recruited based on voluntary participation in the study. Their involvement required prior authorization from the local education authorities and the head teachers of the schools. Teachers were given no remuneration for having being observed. These teachers taught in schools which had already participated in previous stages of the research (with questionnaires) or which had shown interest in joining the observation traineeship developed to assume the observation stage.

Almost two-thirds of the teachers (214; 62.2%) were women, 130 (37.8%) men. About a quarter (25.9%) taught languages, a quarter (25%) science and applied science, 18.3% social sciences, 17.4% vocational education and training subjects, 8.7% cultural and artistic education and 4.7% physical education. Teaching experience ranged from less than five years (n = 34) to over 30 years (n = 42). 215 teachers (62.5%) taught in lower secondary education, 66 (19.2%) in higher secondary education, and 63 (18.3%) in vocational education and training. Teachers worked in classes ranging in size from 2 to 35 students (M = 15.9; SD = 6.4). Interesting internal differences in class size were found in the different educational stages: lower secondary

education (*M* = 17.6; *SD* = 5.9); higher secondary education (*M* = 15.5; *SD* = 7); and vocational education and training (*M* = 10.9; *SD* = 4.9). All the students present in the class group of the volunteer observed teachers stayed during the observation process. There were few students with special educational needs in the observed classes (just 3.7% of the total number of students present in the classrooms during observations). Due to confidentiality matters, we did not record the classes or collect any individual information concerning students. Nevertheless, we did collect general information given by teachers about the general socioeconomic status (SES) and achievement of the groups, which were medium on average.

Instruments

The ICALT observation instrument, validated in several countries, was used to gather information. The original instrument (based on Van de Grift, 2007 version) had been created in English so a back translation into Spanish following the procedure indicated by Hambleton et al. (2005) was needed. With the intention to obtain a homogenous information in all the participants in the project, the Spanish research team selected the same instrument used in the other countries. Moreover, its previous use in so many countries gave more confidence in the success of the research purposes.

The instrument consisted of 35 items grouped into seven domains: "safe and stimulating learning climate" (4 items) (e.g. the teacher maintains a relaxed atmosphere), "efficient classroom management" (4 items) (e.g. the teacher uses the time for learning efficiently),"clarity of instruction" (7 items) (e.g. the teacher teaches in well-structured manner), "activating teaching" (7 items) (e.g. the teacher stimulates the building of self-confidence in weaker leaners), "teaching - learning strategies" (6 items) (e.g. the teacher asks learners to reflect on practical strategies), "differentiation" (4 items) (e.g. the teacher offers weaker learners extra study and instruction time) and "student engagement" (3 items) (e.g. learners take an active approach to learning).

The observers recorded scores about effective teaching practices using a rating scale which ranged from 1 ("mostly weak") to 4 ("mostly strong") in the indicators. On the other hand, in the examples of good practice, answers varied from 0 ("no, I have not observed this") to 1 ("yes, I have observed this"). Therefore, for each high indicator, various examples of good practice (observed items) were given so that observers could establish a factual basis for the score (*e.g. Indicator: The teacher promotes learners' self-confidence. Examples of good practice: the teacher gives positive feedback on questions and remarks from leaners; the teacher compliments learners on their work; the teacher acknowledges the contributions that learners make*). Scores 1-2 indicated low quality teaching practices (indicators in which less than 50% of the examples of good practice were observed), whereas scores 3-4 represented high quality (indicators in which more than 50% of the examples of good practice were observed). The instrument also gave the option to include perceptions, worries, reflections or details observed during the session by the trained observers.

Subjects were categorized in six different groups: language; social science; science and applied sciences, physical education; cultural and artistic education and, finally, vocational education and training (VET) subjects.

Observed teachers' teaching experience considered 5 categories: less than 2 years of experience, 3-9 years of experience, 10-19 years of experience, 20-29 years of experience and more than 30 years of experience.

This instrument provided an interesting opportunity to analyse verbal interactions which represented teachers and students' experiences and which constituted a key point to understand all the personal transformations which took place in the process of teaching and learning. It also provided a framework to reflect about teachers' own practice, how did they cope with complex decisions regarding their teaching and how could they improve this process (Pérez, 2013). The ICALT observation instrument might also help teachers to reflect about their teaching that is,

the systematic and planned activities which promoted the progressive sharing of broader meanings regarding school content between students and their teachers (Coll, 1988).

Procedure

To ensure that observations were sufficiently objective and followed the selected theoretical framework, only trained observers carried out the fieldwork. The traineeship was developed in the University of Oviedo directed by some of the Dutch leaders of the ICALT project. Trainees were organised in various sessions in order to avoid over-large groups. Secondary education teachers and the entire research team (54 observers) participated in the observation training. The training involved broad information about the theoretical/methodological basis of the project and about the observation instrument. In addition, the training included an in–depth explanation of how to codify ratings using two 20-minute videos of secondary education lessons. Levels of observer agreement were calculated and any confusing items were discussed. Cut–off criteria for acceptable consensus among observers was set at ≥ 0.70.

47 trained teachers finally developed the observations. 37 of them were women (79%) and 10 men (21%). 4% of the observers had less than five years of teaching experience; 2% between 3-5 years of experience; 28% between 6-15 years of experience and, finally, 66% more than 15 years of teaching experience.

The observations took place in ordinary classes (excluding, for example, classes in which students were doing written tests) and lasted for approximately 50-55 minutes. Teachers' behaviour was recorded in real-time and the observation instrument was the same for all the disciplines.

Data Analysis

Preliminary analyses to test the reliability and validity of the observation instrument were performed. Firstly, a Confirmatory Factor Analysis (CFA) was run to confirm the original structure (Van de Grift et al., 2014) using MPLUS 7.3 software. Maximum likelihood, was

selected as estimation method. The following measures were included: the Chi-Square test of significance ($\chi 2$), the Tucker Lewis index–non normed fit index (TLI), the Comparative Fit Index (CFI), Steiger's Root Mean Square Error of Approximation (RMSEA) and the Standardized Root Mean Square Residual (SRMR).

In order to analyse the differential functioning of each of the items instead of the traditional Rash model the paradigm of Item Response Theory was employed, using IRTPro Software 4. Consequently, values 1-2 were recoded to 0 (not perceived) and values 3-4 to 1 (perceived) (Noben et al., 2020). The difficulty level, the discrimination and fit indexes were also run.

Cronbach's alpha was used to assess the internal consistency of the scale, and correlational analyses were also performed to examine its relationship with "student academic engagement" as the criterion.

Descriptive analyses were conducted to answer the first research question. To answer the second one, an analysis of variance was performed. The "student engagement" domain was considered as criterion variable and teachers teaching behaviour standards as predictor variables. In addition, the value of eta was measured in order to determine the effect of teacher skills on "student engagement". It should be noted that a $\eta 2$ value of about .01 indicated a small effect size; between .06 and .10 a medium effect size; and over .14 a high effect size (Coe & Merino, 2003; Fritz & MacKinnon, 2007). When findings revealed significant statistical differences within the criterion variable, post hoc analyses of difference were also carried out. A stepwise regression was performed to identify which teaching skill domains were seen as more important when trying to promote "student engagement".

Results

Psychometric quality of the observation instrument

Confirmatory factor analysis exposed good fit of the model, $\chi^2 = 918.56$, df = 518, $p = .000$; TLI = 0.90, CFI = 0.91, RMSEA = .047 [.04, .05], SRMR = .07. Table 1 shows the descriptive

statistics for each of the items. In our sample, Cronbach´s alpha was .93. The reliability scores of the effective teaching behaviour domains were: "safe and stimulating learning climate" (.77), "efficient classroom management" (.78), "clarity of instruction" (.81), "activating teaching" (.72), "teaching–learning strategies" (.86), "differentiation" (.72) and "student engagement" (.88) indicating that all domains were internally consistent.

Table 1.

The differential item functioning showed a good fit: S-χ^2 higher than $p < .01$ (Edelen & Reeve, 2007) (Table 2). Data revealed that those items related with "safe and stimulating learning climate" and "efficient classroom management" were more frequently observed whereas items which imply a higher level of difficulty (e.g. items inside "differentiation" domain) were rarely observed. Focusing on those items which implied a lower level of difficulty, it can be said that item 7 "The teacher provides effective classroom management" and item 9 "The teacher presents and explains the subject material in a clear manner" were the most frequently observed ones. They also obtained the best discrimination indexes.

Table 2.

Table 3 illustrates the instrument validity in terms of mean inter-scale correlations, showing the correlation between the six domains and the criterion variable: "student academic engagement".

Table 3

Positive, significant relationships between the six domains and "student academic engagement" were found: "safe and stimulating learning climate" ($r = 0.32$, $p < .01$), "efficient classroom management" ($r = 0.46$, $p < .01$), "clarity of instruction" ($r = 0.48$, $p < .01$), "activating teaching" ($r = 0.51$, $p < .01$), "teaching–learning strategies" ($r = 0.45$, $p < .01$) and "differentiation" ($r = 0.31$, $p < .01$). The "safe and stimulating learning climate" and

"differentiation" domains exhibited the weakest relationship with "student engagement" whereas "activating teaching" showed the strongest one.

Mean inter-scale correlations showed that although a certain degree of overlap could be deduced by the values in some domains (e.g. "activating teaching" and "clarity of instruction"), the scales did seem to measure different teaching skills satisfactorily: "safe and stimulating learning climate" (.41), "efficient classroom management" (.44), "clarity of instruction" (.58), "activating teaching" (.60), "teaching-learning strategies" (.43) and "differentiation" (.37).

General level of observed teachers' teaching behaviour in Spain

Several categories were established to classify the scores for effective teaching behaviour: unsatisfactory (1-2), satisfactory (2.01 - 3), good (3.01- 3.5) and excellent (3.51- 4) (Authors, 2017a; Authors, 2015a). Scores were converted to percentages to clarify differences in observations.

Table 4

In general, "safe and stimulating learning climate", "efficient classroom management" and "clarity of instruction" were perceived as moderately strong. In the following notes we can clearly understand how do teachers develop these three domains:

- *Students are seated in a circle, with biscuits and drinks (tea and chocolate). All students represent a poem* [it is a group of 10 students] *they talk about the authors' life and even characterize themselves according to the historical context* (Female observer, more than 15 years of teaching experience, observing a female lower secondary education Language teacher)
- *The learning climate is quite good, there is peace, work and authority* (Female observer, between 6-15 years of teaching experience, observing a male higher secondary education English teacher)

- *Instructions are clearly understood by students* (Male observer, between 6-15 years of teaching experience, observing a lower secondary education female Geography teacher)
- *Students know and respect the rules without needing the teacher's intervention. The previous work developed during months to make these rues clear, can easily be perceived* (Female observer, more than 15 years of teaching experience, observing a male lower secondary education Physical Education teacher)
- *The tables and chair have been moved into a "U" shape to enhance collective learning during some activities* (Female observer, between 3-5 years of teaching experience, observing a female higher secondary education Biology teacher)

Nevertheless, one observer also pointed out the need to improve in one of these most basic domains, "safe and stimulating learning climate":

- *Students volume is too loud. They use a colloquial and inappropriate vocabulary, but the teacher takes no action* (Female observer, between 3-5 years of teaching experience, observing a female lower secondary education Chemistry teacher)

On the other hand, "activating teaching", "teaching-learning strategies" and "differentiation" were perceived as moderately weak (Table 4). This pattern suggested that, on average, teachers were seen by observers as better performers in the more basic effective teaching behaviours. Observers reflected about this fact in their "open" notes:

- *Three female students in the classroom have never been asked by the teacher during the classroom. None of them speaks or makes any questions, it seems that they are invisible for the teacher* (Female observer, more than 15 years of teaching experience, observing a male VET Technology teacher)
- *When daily problems are mentioned to stimulate reflection, some students find it difficult to connect the task with their social context. Some of them say that these examples are far from their real life* (Male observer, between 6-15 years of teaching experience, observing a male lower secondary education Geography teacher)

It should be noted that the majority of teachers (40.7% and 41.6%) demonstrated excellent performance in "safe and stimulating learning climate" and "classroom management" while the majority (45.1%) still lack mastery of "differentiation" behaviours, demonstrating unsatisfactory standards. Although as we have previously mentioned there were few students with special educational needs, observers pointed out teachers' behaviours concerning "efficient classroom management" and "safe and stimulating learning climate" which affected groups with this kind of students:

- *The group is formed by students with important learning difficulties (…) Two of them have been expelled from the school due to their bad behaviour. The teacher uses positive reinforcements very frequently* (Female observer, more than 15 years of teaching experience, observing a female lower secondary education English teacher)
- *It is a numerous group. There are important differences between students: there is a group with learning difficulties and another one which has adopted a very passive role. The firs ones, receive support of other teachers and the seconds do not have an easy solution. The diversity observed inside the classroom would require more support, additional teachers and other methodological measures* (Male observer, between 6-15 years of teaching experience, observing a male lower secondary education History teacher)

On the other hand, 38.9%, 52.9% and 48.8% of the sample exhibited satisfactory behaviour in "clarity of instruction", "activating teaching" and "teaching-learning strategies". These observers clearly perceived behaviours related to these domains during the session:

- *Students are organised, know what to do all the time* (Female observer, more than 15 years of teaching experience, observing a female higher secondary education Mathematics teacher)
- *It is a time scheduled lesson. The teacher uses an alarm* (Male observer, less than 3 years of teaching experience, observing a male lower secondary education Geography teacher)
- *The teacher uses TICs, gives a magisterial session and continuously asks students* (Female observer, more than 15 years of teaching experience, observing a female History teacher)

Hence, we can conclude that there is still room for improvement particularly in the less basic domains which imply more complex processes and a greater control of classroom dynamics. It also requires a profound knowledge about the students and the pedagogical requirements derived from their diversity.

Regarding subjects, differences were focused on "safe and stimulating learning climate" (F = 2.78, p = .02, η^2 = .04). Post hoc analysis revealed these differences between language subjects (M = 3.52, SD = 0.50) and VET subjects (M = 3.02, SD = 0.77) (d = -0. 496, p = .000); social sciences (M = 3.51, SD = 0.50) and VET subjects (d = -0. 479, p = .001). On the other hand, no statistical significant differences were found regarding teachers' teaching experience.

Finally, differences according school ownership were also analysed. Private school teachers developed more "activating teaching" strategies than those teachers from public ones (M = 3.1, SD = 0.52; vs M = 2.89, SD = 0.53; F = 5.78, p = .02, η^2 = .02). This finding was similar in "teaching-learning strategies" (private school teachers: M = 2.93, SD = 0.71; public school teacher: M = 2.60, SD = 0.71; F = 7.14, p = .01, η^2 = .02). Likewise, private school "students' engagement" was higher than public school students one (M = 3.40, SD = .67; vs M = 3.15, SD =.72; F = 4.63, p = .03, η^2 = .02).

Influence of the standard of teachers' teaching behaviour on student engagement

An analysis of the parametric requirements was performed in order to test the adequacy of the analysis of variance. All the variables followed a normal distribution; skewness and kurtosis values were lower than 1 in absolute values and the equality of variance was verified by Levene's test. Because some criterion variables did not show equality of variance in the predictor variables, Dunnet's C test was selected for the post hoc analysis of variance.

In terms of engagement (Table 5), findings showed significant statistical differences in the six domains with effect sizes between 22% and 8%. The highest effect sizes were found in "clarity of instruction" and "activating teaching" (22%), followed by "efficient classroom

management" (19%) and "teaching-learning strategies" (18%). The lowest effect sizes were found in "safe and stimulating learning climate" (9%) and "differentiation" (8%) which showed medium values. This observer gave her testimony about the difficulties found by some students to engage with the classroom:

- *Students show great difficulties to follow the lesson, they cannot concentrate. The lesson is very mechanic* (Female observer, more than 15 years of teaching experience, observing a female lower secondary education Music teacher)

Table 5.

Post hoc analysis about "safe and stimulating learning climate" exposed that differences were focused on teachers with excellent skills and the remaining categories (Dunnet's C test is significant at .05). The same differences were found for "efficient classroom management". In the "clarity of instruction" domain differences were focused on teachers with unsatisfactory skills and the remaining categories, and between excellent teachers and the rest. No differences were observed between the middle categories (satisfactory and good). Two domains ("activating teaching" and "teaching-learning strategies") exhibited differences between all categories. Finally, in the "differentiation" domain, Dunnet's C test showed differences between teachers with unsatisfactory skills and the remaining categories.

Once the effect sizes were examined, the aim was to identify which of the six teaching skill domains was pointed out as the most important one for improving "student engagement". To do so, a stepwise regression was performed (Table 6). This information would make us able to identify which teaching learning domains could be key in future teacher training courses. "Activating teaching", "efficient classroom management" and "teaching-learning strategies" were the teaching skills which demonstrated the highest predictive power (together they explained 33% of the variance).

Table 6

Discussion and conclusions

This study has contributed to the validation of a model of teaching behaviour for the Spanish context as a way to improve learning opportunities inside classrooms and helping to provide a framework for instructional improvement grounded in knowledge and information about the teaching process (Kelly et al., 2020). According to Jensen et al. (2019) it has also clarified an explicit framework with clear unit(s) of analysis and theory based conjectures, about how and why specific aspects of teaching affect student learning.

Our results demonstrated that teachers showed good levels of achievement in those basic skills ("safe and stimulating learning climate", "efficient classroom management" and "clarity of instruction") which seem to be a precondition for demonstrating outstanding levels in the other more complex domains ("activating teaching", "teaching-learning strategies" and "differentiation"). Nevertheless, observers have also found some exceptions showing inefficient behaviours, even in these basic domains. Our conclusions confirmed the cumulative order in the levels of difficulty of teachers' teaching behaviours previously found by Authors (2014). In fact, the differential item functioning revealed, in line with Noben et al. (2020), that items such us: "The teacher offers weaker learners extra study and instruction time" and "The teacher adjusts the processing of subject matter to relevant inter-learner differences" were rarely observed and showed the highest level of difficulty.

On the other hand, the observed teachers, on average, demonstrated less outstanding levels in the most complex domains, confirming Hypothesis 1. These findings reinforced similar results found in previous phases of the research but using student ratings (Authors, 2019). This idea has also been reflected in the observers' qualitative statements.

Positive, significant relationships between all the domains and "student academic engagement" were found. "Safe and stimulating learning climate" and "differentiation" were the domains which demonstrated the weakest relationships with "student academic

engagement". On the other hand, "activating teaching" revealed the strongest. Further research is needed about this fact, as only 42.2% of teachers exhibited good/excellent standards in one of the domains which showed a strong relationship with engagement ("activating teaching"). Consistently with previous studies (Bonwell & Eison, 1991; Christensen et al., 2009) this relationship is not surprising, as this domain implies students' awareness of their own learning, critical reflection after activities, and the need to assume an active role as learners which necessarily means that the student is engaged and tries to cope with academic demands. As long as the mere presence of students together will not automatically bring favourable effects, this domain has to do with the fact that teachers have to identify and promote a kind of social organization inside the classroom that makes the attainment of certain educational objectives possible (Cubero et al., 2007).

In the case of "clarity of instruction", which also had a strong relationship with "academic engagement", more than half the teachers (57.3%) demonstrated good/excellent standards. Hence, as suggested by Cardwell (2011) teachers who were concerned about the quality of the education improved their "students' engagement". When the objectives of the lesson were clear, when teachers gave good feedback, when they created learner assignments which stimulate active participation or made sure that all learners knew what to do, observers reported that students paid better attention and participated more actively.

"Efficient classroom management" and "teaching-learning strategies" were also important when the intention was to increase student engagement. Both of them mean that students have internalised a new way of learning, transforming this concept into action: taking the initiative, working independently, and assuming responsibility for their own learning process. As mentioned by Hospel and Galand (2016) this means a balance between teachers' autonomy support and structure. In this context and in line with Danielson (1996) and Oliver & Reschly (2007), the way in which teachers organised classroom processes and their own strategies (e.g.

emphasis on how to simplify complex problems, conscious encouragement to apply what has been learned, encouraging learners to think critically) was revealed as decisive to improve "students' engagement" (Klem & Connell, 2004).

In contrast, as we have previously mentioned, domains such as a "safe and stimulating learning climate" and "differentiation", although exhibiting a medium effect, obtained between 7 and 14 points below the outstanding domains. Perhaps due to the low number of students with special needs in our sample, the "differentiation" domain did not seem to have significantly determined student engagement in general. Further studies should be carried out to analyse the influence of this domain in this particular cohort of students specially when teachers' training does not always consider the heterogeneity of the daily conditions in which teaching takes place (Mercado, 1991). Even though the little presence of this kind of students is a fact, observers identified the attention they require as one of the main challenges teachers had to face with. In the case of "safe and stimulating learning climate", as it was one of the more basic domains, one might conclude that its processes were too general to have a more noticeable influence on "academic engagement".

Differences have been found according to subjects and the ownership of the school. These results reinforce the previously obtained conclusions using teachers and students questionnaires (Authors, 2021a). On the other hand, although in these previous stages differences were also found according teachers' teaching experience, we cannot confirm these data with these results derived from teachers' observation.

Hypothesis 2 has also been confirmed. It is interesting to highlight that in the "activating teaching" and "teaching-learning strategies" domains, differences were found between all categories of teaching skill standards. This fact can indicate that improvements in any of these teaching skills can increase "student engagement", and their chances to improve achievement, reinforcing the conclusions obtained by Maulana and Opdenakker (2013) and Maulana et al.

(2012, 2013). In contrast, differences in the effects of "differentiation" teaching skills were concentrated between unsatisfactory and the other three categories. To find differences in the highest levels other aspects need to be considered. To sum up, when a certain level was reached in this domain, the analysis of potential improvements was more difficult. Our results indicated that if teachers implement "differentiation" strategies at the right time, there would be an improvement in "student engagement". On the other hand, higher standards in "safe and stimulating learning climate" and "efficient classroom management" were needed in order to improve "student engagement". Finally, high standards in "clarity of instruction" will produce an effect on "student engagement". In fact, when teachers did not address the skills associated with this domain, there would be negative effects.

To sum up, our results have suggested that the quality of teachers' teaching skills was an important predictor of "students' academic engagement", which is in line with previous research already mentioned. We can conclude that the most determinant domains in order to improve "student engagement" were: "activating teaching" so that students assumed an active role; "efficient classroom management" which meant among other things that there was supervision of how students performed their tasks and finally, "teaching-learning" strategies which meant that students were taught how to simplify complex problems, how to test problem solutions, use check lists and think critically.

Support seems to be needed in order to help teachers attain sufficiently high standards. On the other hand, giving them a scientific and systematic information that proves that many teachers find it more difficult to achieve outstanding levels in some of these domains, may help them understand that it is not a "personal" problem but the complex consequence of their initial training or the resources stablished by the educational system to support their teaching procedures. This in line with previous research (Danielson, 2012; Jensen et al., 2019; Reddy et al., 2013) which showed that this support may be derived from an efficient feed-back from

observation or student rating results, so that teachers would have the opportunity to reflect about their own practices and be benefited from information about specific effective instructional practices and how best to implement them. Knowledge, and also teachers' one is constructed inside specific social groups and according to certain values (Goodnow, 1996), so a reflection about this experience can provide an accurate framework to obtain this aim. Teachers must understand what happened, analyse their teaching behaviours and reconstruct them taking into consideration theory (Pérez, 2013).

Observation can also help policy-makers improve educational practice by taking decisions based on information obtained from the observation of schools and providing a rational framework for improvement. As we have previously mentioned, this research can be helpful not only for determining progress in classroom practice, but also for guiding initial and continuous teacher training (Kelly et al., 2020). We should not forget that efficient teachers are those who develop a metacognitive control of their own professional knowledge (Carriedo, 1995). Special attention should be paid to the most complex behavioural domains which require long-term interventions, in order to avoid large numbers of teachers never developing these more complex skills. One of the main contributions offered both by this work and the ICALT project is the potentiality offered by the collective work developed in several countries. Hence, this homogenous approach can allow international comparisons (Authors, 2020; Authors 2021b) which may help in the improvement of worldwide teaching processes, while also considering national particularities.

Limitations of the study

This study has several limitations. First, results have been based on one–time observations which can be substantially biased. Our results gave us reliable insight into the specific lessons observed, but more observations would be needed to develop a more complete image of teachers' general teaching proficiency (Van der Lans et al., 2016). A more continuous or

participant observation would make observers perceive the discursive interaction among teachers-students and between students according to the content and the educational function they give to these interactions (Prados & Cubero, 2013). It would also be interesting to have multiple observers visiting the classrooms to give them some anonymity and protection in their ratings (Van der Lans et al., 2016).

The same observation instrument was used for different areas of knowledge and disciplines. Although the instrument focused on general domains which referred to planning, classroom climate or methodologies, tasks which may be observed in almost any teacher, in future research specific instruments should be designed to assure that particularities of all these subjects are sufficiently represented. One observer clearly pointed out the difficulties found when trying to use the instrument in some disciplines (Physical Education, in this case):

- *Items which have been given a rating of 1 or 2 refer to students' thinking or the time needed to understand. These items are difficult to apply in Physical Education, a subject in which action, reflexes and controlled time are determinant* (Female observer, more than 15 years of teaching experience, observing a male lower secondary education Physical Education teacher)

The current observation instrument should be mutually used together with alternative procedures for gathering student and teacher opinions, as long as using more than one source of information (triangulation) might mitigate the weakness of their individual perspectives. Moreover, a more open instrument could also provide the opportunity to analyse and reflect about classroom processes, qualitative achievements, problem solutions or students' success. This alternative approach could give a more profound understanding of teaching-learning process and also promote more reflective analyses based on the strong and weak aspects which have been already detected.

Finally, our sample may also be biased in that it has included teachers who have voluntarily decided to participate in the research. This means that teachers who perceive themselves as

more effective, who feel comfortable in their classes or more motivated, may have been overrepresented in the sample. Also due to this bias, in the observed classes (chosen by teachers between all the groups they taught) there were not many students with special educational needs, for example. Though, it would be desirable in future research to increase the number of participants, the selected groups and randomly sample them.

References

Authors (2014).

Authors (2015a)

Authors (2015b)

Authors (2017a)

Authors (2017b)

Authors (2019)

Authors (2020)

Authors (2021a)

Authors (2021b)

Blaich, C., Wise, K., Pascarella, E. T., & Roksa, J. (2016). Instructional clarity and organization: it's not new or fancy, but it matters. *Change: The Magazine of Higher Learning*, *48*(4), 6-13. https://dx.doi.org/10.1080/00091383.2016.1198142

Bonwell, C. & Eison, J. A. (1991). *Active Learning: Creating Excitement in the Classroom. ASHE-ERIC Higher Education Reports No1*. The George Washington University, School of Education and Higher Education. Briole, S. & Maurin, E. (2019). Does evaluating teachers make a difference? *IZA Discussion Papers*, *123017*, 1 – 39. https://10.13140/RG.2.2.15662.33604

Briole, S. & Maurin, E. (2019). Does evaluating teachers make the difference? *Iza Discussion Paper Series*, *12307*, 1 – 38.

Cardwell, M.E. (2011). Patterns of relationships between teacher engagement and student engagement. *Education Doctoral*, *49,* 1 – 125.

Carriedo, N. (1995). Hacia la contextualización: la enseñanza de estrategias de comprensión de las ideas principales en el aula. *Comunicación, Lenguaje y Educación, 28,* 123-134.

Christensen, H., Vigild, M., Thomsen, E., Szabo, P., & Horsewell, A. (2009). *Activating teaching methods, studying responses and learning*. Arctic Technology Centre & Learning Lab DTU.

Coe, R., & Merino, C. (2003). Magnitud del efecto; Una guía para investigadores y usuarios. *Revista de Psicologıa de la PUCP, 21*, 146–177.

Cole, M. (1984). La zona de desarrollo próximo: donde cultura y conocimiento se generan mutuamente. *Infancia y Aprendizaje, 25,* 3-17.

Coll, C. (1984). Estructura grupal, interacción entre alumnos y aprendizaje escolar. *Infancia y Aprendizaje, 27/28,* 119-138.

Coll, C. (1988). Significado y sentido en el aprendizaje escolar. Reflexiones en torno al concepto de aprendizaje significativo. *Infancia y Aprendizaje, 41*, 131-142.

Connell, J. P., & Wellborn, J. G. (1991). Competence, autonomy, and relatedness: A motivational analysis of self-system processes. In M. R. Gunnar & L. A. Sroufe (Eds.), *Minnesota Symposia on Child Psychology. Self-processes and development* (pp. 43-77). Lawrence Erlbaum Associates.

Consejo Escolar de Estado & Ministerio de Educación y Formación Profesional (2018). *Informe 2018 sobre el estado del sistema educativo.* Ministerio de Educación y Formación Profesional. http://www.educacionyfp.gob.es/educacion/mc/cee/publicaciones/informes-del-sistema-educativo/informe-2018.html

Cubero, R., Cubero, M., Santamaría, A., Saavedra, J. & Yossef, J.J. (2007). Aprendizaje y psicología histórico-cultural. Aportaciones de una perspectiva social del aula. *Investigación en la escuela, 62*, 5-16. https://doi.org/10.12795/IE.2007.i62.01

Danielson, C. (1996). *Enhancing Professional Practice: A Framework for Teaching*. Association for Supervision and Curriculum Development (ASCD).

Danielson, C. (2012). Observing classroom practice. *Educational Leadership, 70* (3), 32 – 37.

Danielson, C. (2013). Rubrics from the framework for teaching evaluation instrument. https://www.ode.state.or.us/wma/teachlearn/commoncore/danielson-2013-rubric-only.pdf

De Jager, T. (2011). Guidelines to assist the implementation of differentiated learning activities in South African secondary schools. *International Journal of Inclusive Education*, *17*(1), 80-94. https://doi.org/10.1080/13603116.2011.580465

De Jong, R. & Westerhof, K.J. (2001). The quality of student ratings of teacher behaviour. *Learning Environments Research*, *4*, 51–85. https://dx.doi.org/10.1023/A:1011402608575

Edelen M.O. & Reeve B.B. (2007). Applying item response theory (IRT) modeling to questionnaire development, evaluation, and refinement. *Quality of Life Research, 16* (1), 5-18. https// 10.1007/s11136-007-9198-0

Emmer, E. T. & Stough, L.M. (2001). Classroom management: a critical part of educational psychology, with implications for teacher education. *Educational Psychologist, 36*(2), 103-112. https://dx.doi.org/10.1207/S15326985EP3602_5

Finn, J. D. (1993). *School engagement and students at risk*. National Center for Educational Statistics.

Fredricks, J., Blumenfeld, P., & Paris, A. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research*, *74* (1), 59–109. https://dx.doi.org/10.3102/00346543074001059

Fritz, M., & MacKinnon, D. (2007). Required sample size to detect the mediated effect. *Psychological Science, 12*, 233–238. https://dx.doi.org/10.1111/j.14678-9280.2007.01882.x

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. https://dx.doi.org/10.1037/0022-0663.95.1.148

Gage, N. A., Scott, T., Hirn, R. & MacSuga-Gage, A.S. (2018). The relationship between teachers' implementation of classroom management practices and student behavior in elementary school. *Behavioral Disorders, 43*(2), 302 – 315. https://doi.org/10.1177/0198742917714809

García-Garrido, J.L. (2002). La nueva Ley de Calidad: reflexiones en la perspectiva internacional. *Revista de Educación, 329*, 23 - 38.

Gargallo, B., Pérez-Pérez, C., García-García, F. J., Giménez, J. A. y Portillo, N. (2020). La competencia aprender a aprender en la universidad: propuesta de modelo teórico. *Educación XX1*, *23*(1), 19-44. https://doi.org/10.5944/educXX1.23367.

Gispert, I. & Onrubia, J. (1997). Analizando la práctica educativa con herramientas socio-culturales: traspaso del control y aprendizaje en situaciones de aula. *Cultura y Educación, 9* (2/3), 105-115. https://doi.org/10.1174/113564097761403526

Goodnow, J.J. (1996). Acceptable ignorance, negotiable disagreement: alternative reviews of learning. In D.R. Olson & N. Torrance (Eds.), *The handbook of education and human development* (pp. 345-367). Blackwell Publishers.

Hamre, B. K., Pianta, R. C., Downer, J. T., DeCoster, J., Mashburn, A. J., Jones, S. M., Brown, J. L., Cappella, E., Atkins, M., Rivers, S. E., Brackett, M. A., & Hamagami, A. (2013). Teaching through interactions: Testing a developmental framework of teacher effectiveness

in over 4,000 classrooms. *The Elementary School Journal, 113,* 461–487. https://doi.org/10.1086/669616

Hambleton, R.K., Merenda, P., & Spielberger, C. (eds.) (2005). *Adapting educational and psychological tests for cross-cultural assessment*. Lawrence Erlbaum Publishers.

Hanushek, E. A. (2016). What matters for student achievement. *Education Next, 16*(2), 19-26.

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*. https://research.acer.edu.au/research_conference_2003/4/

Hospel, V. & Galand, B. (2016). Are both classroom autonomy support and structure equally important for students' engagement? A multilevel analysis. *Learning and Instruction, 41*, 1-10. https://doi.org/10.1016/j.learninstruc.2015.09.001.

Howe, C., Hennessy, S., Mercer, N., Vrikki, M., & Wheatley, L. (2019). Teacher–Student Dialogue During Classroom Teaching: Does It Really Impact on Student Outcomes?. *Journal of the Learning Sciences, 28* (4-5), 462-512. https://doi.org/10.1080/10508406.2019.1573730

Hoyt, D. P. & Pallet, W. H. (1999). Appraising teaching effectiveness: beyond student ratings. *Idea Paper, 36*, 1-14.

Jensen, B., Grajeda, S., & Haertel, E. (2018). Measuring cultural dimensions of classroom interactions. *Educational Assessment, 23*(4), 250-276.https://doi.org/10.1080/10627197.2018.1515010

Jensen, B., Wallace, T.L., Steinberg, M.P., Gabriel, R.E., Dietiker, L., Davis, D.S., Kelcey, B., Minor, E,, Halpin, P. & Rui, N. (2019). Complexity and scale in teaching effectiveness research: Reflections from the MET Study. *Education Policy Analysis Archives, 27*(7), 1 – 17. . http://dx.doi.org/10.14507/epaa.27.3923

John –Steiner, V. & Mahn, H. (1996). Sociocultural approaches to learning and development: a Vygostkyan framework. *Educational Psychologist, 31(*3/4), 191-206.

Kelly, S., Bringe, R., Aucejo, E., & Fruehwirth, J. (2020). Using global observation protocols to inform research on teaching effectiveness and school improvement: Strengths and emerging limitations. *Education Policy Analysis Archives, 28*(62), 1 – 29..

Klem, A. M. & Connell, J. O. (2004). Relationships matter: linking teacher support to student engagement and achievement. *Journal of School Health, 74*(7), 262-273. https://doi.org/10.1111/j.1746-1561.2004.tb08283.x

Kugel, P. (1993). How professors develop as teachers. *Studies in Higher Education,* 18(3), 315 – 328.

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. https://dx.doi.org/10.1016/j.tate.2013.07.010

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. https://dx.doi.org/10.1016/j.tate.2008.06.001

Martínez Usarralde, M.J. (2015). Spain. In W. Hörner, H. Döbert., L.R. Reuter, & B. Von Kopp (Eds.), *The Education Systems of Europe. Second edition* (pp. 759 – 777). Springer.

Maulana, R., & Opdenakker, M.-C. (2013). Teachers' interpersonal involvement as a predictor of students' academic motivation among Indonesian secondary school students: A multilevel growth curve analysis. *The Asia-Pacific Education Researcher,* 23, 591-603. http://dx.doi.org/10.1007/s40299-013-0132-7

Maulana, R., Opdenakker, M.-C., & Bosker, R. (2013). Teacher–student interpersonal relationships do change and affect academic motivation: A multilevel growth curve modelling. *British Journal of Educational Psychology*. 84, 459-482.http://dx.doi.org/10.1111/bjep.12031

Maulana, R., Opdenakker, M.-C., Stroet, K., & Bosker, R. (2012). Observed lesson structure during the first year of secondary education: Exploration of changeand link with academic engagement. Teaching and Teacher Education, 28(6), 835–85. https://doi.org/10.1016/j.tate.2012.03.005

Mercado, R. (1991). Los saberes docentes en el trabajo cotidiano de los maestros. *Infancia y Aprendizaje, 55*, 59-72.

Ministerio de Educación, Cultura y Deporte de España (2016). *PISA 2015. Programa para la evaluación internacional de los alumnos. Informe español.* Ministerio de Educacion, Cultura y Deporte.

Muijs, D. & Reynolds, D. (2000). School effectiveness and teacher effectiveness in mathematics: some preliminary findings from the evaluation of the Mathematics Enhancement Programme (Primary*). School Effectiveness and School Improvement*, *11* (3), 273 – 303. https://dx.doi.org/10.1076/0924-3453(200009)11:3;1-G;FT273

Noben, I., Maulana, R., Deinum, J.F. & Adriaan-Hofman,W.H. (2020). Measuring university teachers' teaching quality: a Rasch modelling approach. *Learning Environments Research*, *24*, 87 – 107. https://doi.org/10.1007/s10984-020-09319-w.

Oliver, R. M. & Reschly, D. J. (2007). *Effective classroom management: teacher preparation and professional development*. National Comprehensive center for Teacher Quality. https://files.eric.ed.gov/fulltext/ED543769.pdf

Opdenakker, M. V. & Minnaert, A. (2011). Relationships between learning environment characteristics and academic engagement. *Psychological Reports, 109*(1), 259-284. https://dx.doi.org/10.2466/09.10.11.PR0.109.4.259-284

Organisation for Economic Cooperation and Development (2018). *PISA 2015: PISA results in focus*. https://www.oecd.org/pisa/pisa-2015-results-in-focus.pdf

Pérez, M. (2013). La actitud indagadora del profesor: un proceso para desarrollar competencias en los docentes. *Profesorado. Revista de curriculum y formación del profesorado, 17*(3), 57-72.

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. https://dx.doi.org/10.3102/0013189X09332374

Puelles, M. (2008). Las grandes leyes educativas de los últimos doscientos años. *Participación educativa, 7*, 7 - 15.

Prado, M. & Cubero, M. (2013). Reflexionando acerca de cómo estudiar el proceso de enseñanza-aprendizaje en el contexto natural del aula universitaria. *Cultura y Educación, 25*(3), 273-284. https://doi.org/10.1174/113564013807749696

Reddy, L. A., & Dudek, C.M. (2014). Teacher progress monitoring of instructional and behavioral management practices: An evidence-based approach to improving classroom practices. *International Journal of School & Educational Psychology*, *2*, 71–84. https://doi.org/10.1080/21683603.2013.876951

Reddy, L.A., Fabiano, G., Dudek, C.M. & Hsu. L. (2013). Development and construct validity of the classroom strategies scale-observer form. *School Psychology Quarterly, 28* (4), 317–341. https://doi.org/10.1037/spq0000043

Reyes, M.R., Brackett, M.A., Rivers, S.E., White, M. & Salovey, P. (2012). Classroom emotional climate, student engagement and academic achievement. *Journal of Educational Psychology*, 104, 700 -712. https://dx.doi.org/10.1037/a0027268

Sammons, P., & Bakkum, L. (2011). Effective schools, equity and teacher effectiveness: a review to the literature. *Profesorado, Revista de Curriculum y Formación del Profesorado, 15*(3), 9-26.

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. https://dx.doi.org/10.1037/0022-0663.85.4.571

Skinner, E., Furrer, C., Marchand, G., & Kindermann, T. (2008). Engagement and disaffection in the classroom: Part of a larger motivational dynamic? *Journal of Educational Psychology, 100*(4), 765-781. https://10.1037/a0012840

Skinner, E.A., Kindermann, T.A., & Furrer, C.J. (2008). A motivational perspective on engagement and disaffection: Conceptualization and assessment of children's' behavioural and emotional participation in academic activities in the classroom. *Educational and Psychological Measurement, 69,* 493-525. http://10-1177/0013164408323233

Tuytens, M., & Devos, G. (2014). How to activate teachers through teacher evaluation. School *Effectiveness and School Improvement, 25*(4), 509-530, https://10.1080/09243453.2013.842601

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. https://dx.doi.org/ 10.1080/00131880701369651

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. https://dx.doi.org/10.1080/09243453.2013.794845

Van de Grift, W.; Helm – 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. http://dx.doi.org/10.1016/j.stueduc.2014.09.003

Van de Grift, W., & Lam, J. (1998). Het didactisch handelen in het basisonderwijs. Tijdschrift voor Onderwijsresearch, *23*(2), 224-241.

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. https://dx.doi.org/10.1111/emip.12078

Van der Lans, R. M., Van de Grift, W., Van Veen, K., & Fokkens-Bruinsma, M. (2016). Once is not enough: establishing reliability criteria for feedback and evaluation decision based on classroom observations. *Studies in Educational Evaluation, 50*, 88-95. https://dx.doi.org/10.1016/j.stueduc.2016.08.001

Viñao, A. (2016). La Ley Orgánica de Mejora de la Calidad Educativa (LOMCE) de 2013: ¿una reforma más? *Historia y Memoria de la Educación*, *3*, 137-170.

Vygostky, L. S. (1979). *Pensamiento y lenguaje*. La Pléyade.

Vygostky, L. S. (1984). Aprendizaje y desarrollo intelectual en la edad escolar. *Infancia y Aprendizaje, 27/28,* 105-116.

Wang, M. C., Haertel, G. D., & Walberg, H. J. (1993). Toward a Knowledge base for school learning. *Review of Educational Research*, *63*(3), 249-294. https://dx.doi.org/10.3102/00346543063003249

White, M. C. (2018). Rater performance standards for classroom observation measures. *Educational Researcher, 47*, 492–501. https://doi.org/10.3102/0013189X18785623

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. https://dx.doi.org/10.1111/j.1741-3729.2007.00442.x

Table 1.
Descriptive statistics in ICALT observation instrument

Domain	Item	Item Content	Mean	Variance	Loadings
		The teacher			
LC	1	… shows respect for learners in their behaviour and language.	3.71	0.32	.58
LC	2	… maintains a relaxed atmosphere.	3.55	0.47	.70
LC	3	… promotes learners' self-confidence.	3.20	0.76	.74
LC	4	… fosters mutual respect.	3.01	0.89	.66
EM	5	… ensures the lesson proceeds in an orderly manner.	3.34	0.61	.84
EM	6	… monitors to ensure learners carry out activities in the appropriate manner.	3.19	0.67	.54
EM	7	… provides effective classroom management.	3.47	0.52	.77
EM	8	… uses the time for learning efficiently .	3.46	0.55	.66
CI	9	… presents and explains the subject material in a clear manner.	3.36	0.57	.69
CI	10	… gives feedback to learners.	3.29	0.69	.57
CI	11	… engages all learners in the lesson.	3.13	0.64	.69
CI	12	… during the presentation stage, checks whether learners have understood the subject material.	3.18	0.81	.62
CI	13	… encourages learners to do their best.	2.54	1.05	.49
CI	14	… teaches in a well-structured manner.	3.23	0.67	.65
CI	15	… gives a clear explanation of how to use didactic aids and how to carry out assignments.	3.17	0.70	.60
AT	16	… offers activities and work forms that stimulate learners to take an active approach.	3.00	0.69	.49
AT	17	… stimulates the building of self-confidence in weaker leaners.	2.26	1.09	.45
AT	18	… stimulates learners to think about solutions.	2.76	0.66	.60
AT	19	… asks questions which stimulate learners to reflect.	3.05	0.63	.66
AT	20	… lets learners think aloud.	3.38	0.65	.50
AT	21	… gives interactive instructions.	3.28	0.61	.48
AT	22	… clearly specifies the lesson aims at the start of the lesson.	3.01	1.02	.51
DI	23	… evaluates whether the lesson aims have been reached.	2.46	1.15	.87
DI	24	… offers weaker learners extra study and instruction time.	1.90	0.98	.67
DI	25	… adjusts instructions to relevant inter-learner differences.	2.45	1.08	.67
DI	26	… adjusts the processing of subject matter to relevant inter-learner differences.	2.05	0.87	.78
TL	27	… teaches learners how to simplify complex problems.	2.64	1.18	.76
TL	28	… stimulates the use of control activities.	2.54	1.01	.77
TL	29	… teaches learners to check solutions.	2.63	0.89	.80
TL	30	… stimulates the application of what has been learned.	3.03	0.81	.61
TL	31	… encourages learners to think critically.	2.84	0.88	.55
TL	32	… asks learners to reflect on practical strategies.	2.42	1.01	.73
AE	33	… are fully engaged in the lesson.	3.30	0.60	.85
AE	34	… show that they are interested.	3.23	0.68	.93
AE	35	… take an active approach to learning	3.11	0.63	.75

Table 2.

Analysis of ICALT observation instrument

Dimension	Item	Descriptive Statistics					Fit Index	
		p_1	*b*	*SE*	*a*	*SE*	$S\text{-}\chi^2$	p_2
LC	1	.98	-2,78	0,56	1,78	0,58	3.51	.48
LC	2	.94	-2,38	0,42	1,44	0,37	16.41	.13
EM	8	.90	-1,87	0,26	1,63	0,34	13.34	.58
CI	10	.86	-1,77	0,27	1,29	0,26	21.22	.27
AT	20	.88	-1,72	0,23	1,56	0,31	18.19	.31
EM	7	.92	-1,62	0,17	2,55	0,53	16.23	.18
AT	21	.85	-1,61	0,22	1,48	0,29	17.28	.57
EM	5	.86	-1,6	0,21	1,65	0,32	10.35	.84
CI	9	.88	-1,52	0,17	2,16	0,41	15.26	.43
EM	6	.82	-1,48	0,21	1,34	0,25	18.10	.45
AE	33	.87	-1,41	0,15	2,25	0,42	18.66	.23
CI	14	.83	-1,37	0,17	1,65	0,3	13.52	.76
AE	35	.81	-1,26	0,16	1,62	0,29	14.87	.67
LC	3	.78	-1,23	0,18	1,38	0,24	19.26	.44
CI	11	.81	-1,2	0,14	1,89	0,32	19.12	.21
AE	34	.81	-1,18	0,14	1,95	0,34	15.83	.39
CI	15	.79	-1,17	0,15	1,68	0,29	21.68	.36
CI	12	.80	-1,15	0,14	1,84	0,31	15.59	.55
AT	16	.73	-1,12	0,19	1,12	0,2	20.13	.45
TL	30	,74	-1,1	0,17	1,21	0,22	19.26	.51
AT	19	.79	-1,05	0,12	1,99	0,33	17.17	.44
LC	4	.73	-0,99	0,16	1,3	0,22	21.23	.39
AT	22	.68	-0,86	0,17	1,08	0,19	17.99	.71
TL	31	.67	-0,68	0,12	1,48	0,24	22.92	.19
AT	18	.65	-0,67	0,13	1,28	0,21	15.65	.74
TL	27	.56	-0,25	0,1	1,79	0,27	13.04	.67
TL	29	.54	-0,19	0,11	1,44	0,22	24.11	.15
TL	28	.53	-0,13	0,11	1,37	0,21	19.07	.45
CI	13	.48	0,02	0,11	1,6	0,23	16.19	.51
DI	23	.48	0,02	0,14	0,99	0,17	21.46	.43
DI	25	.47	0,16	0,16	0,85	0,16	21.84	.41
TL	32	.45	0,16	0,11	1,68	0,25	10.85	.76
AT	17	.40	0,45	0,15	1,12	0,18	21.96	.23
DI	26	.27	1,13	0,21	1,07	0,19	16.67	.41
DI	24	.22	1,68	0,33	0,88	0,18	19.55	.24

Note. p_1 = Proportion of correct responses, *b* = Difficulty Level, *a* = Slope (Discrimination Index), $S\text{-}\chi^2$ = fit index, *SE* = Standard Error, p_2 = Significance level of $S\text{-}\chi^2$ statistic.

Table 3

Correlations between teachers' teaching behavior and student academic engagement

	LC	EM	CI	AT	TL	DI
Safe a stimulating learning climate (LC)						
Efficient classroom management (EM)	.39 **					
Clarity of instruction (CI)	.57 **	.66 **				
Activating teaching (AT)	.59 **	.52 **	.76 **			
Teaching-learning strategies (TL)	.29 **	.34 **	.51 **	.62 **		
Differentiation (DI)	.24**	.30 **	.44 **	.51 **	.38**	
Academic engagement (AE)	.32 **	.46 **	.48 **	.51 **	.45 **	.31 **
M	*M*	3.36	3.36	3.12	3.96	2.68
SD	0.60	0.59	0.58	0.53	0.75	0.75

* $p < .05$ ** $p < .01$ *** $p < .001$

Table 4

ICALT – domains scores

	Unsatisfactory *n* (%)	Satisfactory *n* (%)	Good *n* (%)	Excellent *n* (%)
Safe and stimulating learning climate	11(3.2%)	94(27.3%)	99 (28.8%)	140 (40.7)
Efficient classroom management	11 (3.2%)	104 (30.2%)	86 (25%)	143 (41.6%)
Clarity of instruction	15 (4.4%)	132(38.4%)	99 (28.8%)	98 (28.5%)
Activating teaching	17(4.9%)	182(52.9%)	88 (25.6%)	57(16.6%)
Teaching - learning strategies	78 (22.7%)	168(48.8%)	48(14%)	50(14.5%)
Differentiation	155 (45.1%)	147 (42.7%)	26(7. 6%)	15(4.4%)

Table 5.

Influence of the standard of teachers' teaching skills on student engagement

	Unsatisfactory		Satisfactory		Good		Excellent		F	η^2
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*		
Safe and stimulating learning climate	2.61	1.13	2.97	0.72	3.19	0.66	3.44	0.64	11.61***	.09
Efficient Classroom management	2.15	1.07	2.97	0.71	3.14	0.65	3.52	0.55	26.21***	.19
Clarity of instruction	2.07	1.02	3.04	0.65	3.25	0.64	3.59	0.53	31.22***	.22
Activating teaching	2.29	1.04	3.05	0.65	3.38	0.63	3.75	0.41	31.61***	.22
Teaching-learning strategies	2.81	0.86	3.16	0.60	3.44	0.62	3.79	0.44	25.54***	.18
Differentiation	3.02	0.71	3.31	0.71	3.58	0.54	3.69	0.48	9.73***	.08

Table 6

Predictors of student engagement

	Student engagement								
				Model 2			Model 3		
Variable	Model 1 *B*	95%IC	VIF	*B*	95%CI	VIF	*B*	95%CI	VIF
Constant	1.2	[0.83,1.57]		0.68	[0.27,1.04		0.67	[0.26,1.07]	
Activating teaching	0.68	[0.56,0.80]	1.00	0.50	[0.36,0.64]	1.38	0.34	[0.17,0.50]	1.98
Efficient classroom management				0.32	[0.19,0.44]	1.38	0.31	[0.19,0.43]	1.38
Teaching-learning strategies							0.19	[0.08,0.30]	1.62
R^2	.26			.31			.33		
F	119.27***			75.93***			56.35***		
ΔR^2	.26			.05			.02		
ΔF	119.27			24.43			12.20		

Note = CI = confidence interval. VIF = Variance Inflation Factor

* $p < .05$ ** $p < .01$ *** $p < .001$