



Physical Education teaching competencies assessed with observational instruments in randomised studies: a systematic review

Competencias docentes en Educación Física evaluadas con instrumentos observacionales en estudios aleatorizados: una revisión sistemática

Authors

Jesús Mellado-Berenguer ¹
Manuel Monfort-Pañego ¹

¹ University of València, València, Spain

Corresponding author:
Jesús Mellado-Berenguer
jemebe@alumni.uv.es

Received: 14-06-2025

Accepted: 13-08-2025

How to cite in APA

Mellado-Berenguer, J., & Monfort-Pañego, M. (2025). Physical Education teaching competencies assessed with observational instruments in randomised studies: a systematic review. *Retos*, 72, 410-424. <https://doi.org/10.47197/retos.v72.116705>

Abstract

Introduction: Direct observation has been used as a method to analyse teacher practices in Physical Education for years. Although some systematic reviews have focused on the role of teachers in Physical Education, none of them have centred on the observation instruments used to assess teaching competencies.

Objective: This systematic review aims to analyse the observational instruments used to assess teaching competencies in Physical Education, identify the competencies that they include, and examine observation methods and observer training processes in papers which used a randomised sample.

Methodology: The PRISMA protocol was followed, and the search yielded 783 articles. Inclusion criteria were related to the appearance of the competencies, the use of observational instruments, and the randomization of the sample.

Results: Seven papers exceeded the inclusion criteria. The most common teaching competence was providing feedback to students, but others such as lesson management, student involvement, or the use of informative language were found. The studies used various observational instruments for the evaluation of teaching competencies.

Discussion: The analysis of the observation tools revealed a variety of approaches, none of which attempted to provide a comprehensive analysis of teaching competencies in Physical Education.

Conclusions: We emphasize the need to develop an observational instrument to assess the essential teaching competencies that Physical Education teachers should possess.

Keywords

Observational instruments; Physical Education; systematic review; teacher assessment; teaching competencies.

Resumen

Introducción: La observación directa ha sido utilizada como método para analizar las prácticas docentes en Educación Física durante años. Aunque algunas revisiones sistemáticas se han centrado en el rol de los docentes en Educación Física, ninguna de ellas ha puesto el foco en los instrumentos de observación utilizados para evaluar las competencias docentes.

Objetivo: Esta revisión sistemática pretende analizar los instrumentos observacionales utilizados para evaluar las competencias docentes en Educación Física, identificar las competencias que incluyen y examinar los métodos de observación y los procesos de formación de los observadores, en artículos que utilizan muestras aleatorizadas.

Metodología: Se siguió el protocolo PRISMA y la búsqueda realizada proporcionó un total de 783 artículos. Los criterios de inclusión estuvieron relacionados con la aparición de las competencias en las investigaciones, el uso de instrumentos de observación y con la aleatorización de la muestra.

Resultados: Siete artículos cumplieron con estos criterios. La competencia docente más común fue la capacidad de proporcionar retroalimentación a los estudiantes, pero otras como la gestión de la sesión, la capacidad para involucrar al alumnado o el uso de un lenguaje informativo también fueron encontradas. Los estudios seleccionados utilizaron varios instrumentos de observación para evaluar las competencias docentes.

Discusión: El análisis sobre los instrumentos observacionales utilizados reveló distintos enfoques, pero ninguno de ellos trató de analizar de manera global las competencias docentes en Educación Física.

Conclusiones: Enfatizamos la necesidad de desarrollar un instrumento observacional para evaluar las competencias docentes esenciales que los maestros de Educación Física deberían poseer.

Palabras clave

Competencias docentes; Educación Física; evaluación docente; instrumentos de observación; revisión sistemática.



Introduction

Scientific literature has made great efforts to explain the problem of teaching quality or effectiveness (Maulana et al., 2023). Current educational research has identified teacher quality as one of the essential factors to improve the quality of education in developed countries due to its consistent relationship with student achievement (Kunter et al., 2013). Teachers' behaviours have the potential to motivate students, which is related to positive outcomes, and they have a great influence on students' perceptions of autonomy and competence (Vasconcellos et al., 2020). As Kunter et al. (2013) explained, teachers with strong pedagogical content knowledge (PCK), constructivist beliefs, and enthusiasm for teaching produce higher student achievement. These statements imply that the role of the teacher is of great importance for the achievements of the students and, consequently, it enhances the educational quality. All this evidence has led governments, institutions in different countries (UNESCO, 2015) and researchers to intensify research into teacher education and professional development (PD) (Postholm, 2012) from different fields of knowledge. In Physical Education (PE), research has also been conducted into PD and teacher education (Ko et al., 2006; Lee et al., 2019; Setiawan et al., 2024), taking into account the specificity of its context.

However, should PD programs contribute to teacher quality or to quality teaching? Though both concepts are related, and have a huge impact on positive student outcomes (Berckley, 2019), they are not equivalent. Researchers and policy-makers have identified teacher quality as the most important predictor of student academic success (Fitchett & Heafner, 2018), but the assumption that quality teaching depends entirely on teacher quality is inconsistent, because "even highly qualified teachers are still at the mercy of unreliable circumstances that affect the quality of their teaching" (Kennedy, 2006, p. 19). However, traditional education policy has focused more on teacher characteristics and credentials rather than on quality teaching (Fitchett & Heafner, 2018). Hiebert and Morris (2012) pointed out that increasing teaching is the most productive option in order to improve classroom instruction, i.e., it is more useful to enhance the methods used to interact with students regarding content rather than focusing entirely on teacher talent or qualifications. One important way to improve quality teaching is through PD programs, which should be oriented to small details of practice (e.g., interruptions, material dependence, or student engagement), as these details have a great influence on teaching quality (Kennedy, 2006). As investigations into teachers' instructional behaviours provide exemplars of quality teaching (Fitchett & Heafner, 2018), this study aims to contribute to quality PE teaching through the analysis of the observational instruments that assess PE teaching competencies.

Assessing teaching competencies through observation

The research methodologies used to study quality teaching have been very diverse. Some studies have employed teacher perception as a method to analyse the quality of their teaching (Asun et al., 2020; Martínez-Angulo et al., 2023), others have collected data from various sources (Kawuryan et al., 2021), and some have conducted direct observations in the field (Vattøy & Gamlem, 2020) or through video recordings (Liang, 2015). Thus, the instruments used to assess quality teaching have also varied: surveys and questionnaires, behavioural observations, examinations, interviews, portfolios and recordings. However, the best way to identify and promote good teacher practices is through a combination of student achievement assessment and teacher observation (Rink, 2013). As McKenzie (2010) concludes, direct observation is especially important due to its capacity to generate information about how students behave during lessons, how the lessons are conducted (context), and how teachers manage time.

Direct observation has been used as a method to analyse teacher practices in PE for years (Ko et al., 2006) and it has been employed to investigate various aspects in PE. Curtner-Smith and Todorovich (2002), for example, constructed an instrument to assess teaching competencies that favoured a climate of task involvement in PE lessons. De Souza et al. (2017) combined this method with memory elicitation processes; that is, after the lessons were recorded, the researchers showed them to the participants so that they could express their perceptions and assessments of what had happened. Likewise, direct observation instruments have been applied to measure fidelity in the instructional implementation of different pedagogical methods, such as the teaching of personal and social responsibility (Escartí et al., 2018) or teaching games for understanding (Roberts & Fairclough, 2011).



Although some systematic reviews have focused on the teacher's role in PE (Lander et al., 2016; Villaverde-Caramés et al., 2021), to the extent of our knowledge, this is the only one centred on the observational instruments used and their teaching competencies, which shows the impact of this study due to its originality and novelty.

Objective

The purpose of this systematic review is to analyse the observational instruments used to assess teaching competencies in PE, identify the competencies that they include, and examine observation methods and observer training processes in papers which used a randomised sample.

Method

Search protocol

The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) protocol was followed for this systematic review. The PRISMA statement consists of a 27-point checklist and a four-phase flowchart, and it aims to help authors improve the reporting of systematic reviews (Moher et al., 2009). The search for scientific articles was carried out using the Web of Science (WOS) and ERIC databases. The WOS search included records from the following: Science Citation Index Expanded (SCIE); Social Sciences Citation Index (SSCI); Arts and Humanities Citation Index (AHCI); Current Chemical Reactions; Index Chemicus, ISI Proceedings - Science and Technology (ISTP); ISI Proceedings - Social Science and Humanities Edition (ISSHP); Book Citation Index, Emerging Sources Citation Index; MEDLINE; Scielo Citation Index; Korean Citation Index and the Russian Science Citation Index. The ERIC search included records from the Current Index of Journals in Education and the Resources in Education Index.

The search strategy used for both databases included the following keywords. To delimit the search to the subject area, the term "Physical Education" was used. Since the review focuses on teachers, the terms "teacher" and "pre-service teacher" were included. To find articles referring to quality teaching, we used "professional development", effective*, efficacy* and quality*, applying the symbol "*" to also search for terms belonging to the same lexical family. Finally, the words "peer observation", "systematic observation", "intervention", "field-based" and "field observation" were used to find interventions and observations that provided evidence of quality teaching. The Boolean operators AND and OR were used to combine the different concepts. The search expression was: "Physical Education" and (teacher or "pre-service teacher") and ("professional development" or effective* or efficacy* or quality*) and ("peer observation" or "systematic observation" or intervention or "field-based" or "field observation").

Inclusion criteria

Once the search was conducted, and any repeated papers were discarded, the relevant articles were selected considering the following inclusion criteria. Firstly, PE is a subject that is not taught by specialized teachers in all education systems, so in order to avoid problems of bias in the selection of the study population it was decided that participants had to be PE teachers, or future PE teachers or classroom teachers who teach PE classes. Secondly, due to the consistent relationship between students' achievement and quality teaching (Fitchett & Heafner, 2018), PE teaching competencies in primary or secondary education had to be specified. Thirdly, the idea of what is good practice depends on the context, hence perceptions could be quite different around the world and highly subjective. To avoid perception bias, studies had to collect evidence on effective teaching or quality teaching through observational instruments. Fourthly, in contrast to randomized controlled trials in education, cluster randomization trials reduce the probability of sample contamination bias or peer effects (Goesling, 2019), for this reason cluster randomization was included. Finally, the papers had to be written in English or Spanish.

Data collection process

The final search was conducted during May 2024. One of the researchers was responsible for the final search and for removing any repeated articles. Four papers from previous searches were added because the authors felt they could meet the inclusion criteria, even though they did not appear in the final



search. Only one of them was finally selected. After the search, the inclusion criteria were applied by reading the titles and abstracts of the filtered articles. Papers that clearly failed to meet one or more of the inclusion criteria were discarded.

After this first selection, a double-blind strategy was used in the process. The full texts of the articles were retrieved and carefully read by two researchers, separately, in order to apply the inclusion criteria. The procedure followed consisted of using a double-entry table specifying whether the inclusion criteria were met. In cases where doubt existed, the application of the criteria was argued in the table for subsequent discussion. The selections made individually was contrasted by the researchers in a meeting. In cases of disagreement, the original documents were read by both researchers, together, and the arguments were discussed until a consensus decision was reached.

To record all the data on the fundamental characteristics of the selected studies, a spreadsheet was used to collect the information.

Risk of bias assessment

The articles finally included in the review were subjected to a risk-of-bias assessment using a 13-item scale based on the research of Lander et al. (2016) and Van Sluijs et al. (2007) (see Table 1). Two researchers individually evaluated the studies, assessing compliance with each of the items of the scale. For each item, a "yes" was used in the case of compliance, a "no" in the case of non-compliance and a "?" when it was not specified, unclear or not relevant to the type of research. In their initial assessments, the level of agreement of the two researchers was 97.44%. For those items on which there was no initial agreement, a joint evaluation was carried out and a consensus was reached. Supplementary Material I shows the final ratings given to each article for each item (available at <https://osf.io/438za/files/osfstorage>).

Table 1. Risk of bias checklist

Item	Description
A	Randomization (sequence assignment generation, assignment concealment, and implementation) clearly described and adequately completed.
B	Validated instruments for teacher data collection.
C	Validated instruments for student data collection.
D	Evaluator preparation clearly described.
E	Baseline data presented separately according to treatment group.
F	Participants analysed in the group they were originally in, and not excluded due to treatment noncompliance or loss of data.
G	Intergroup variables analysed in the results (gender, age, years of experience, etc.).
H	Statistical power calculated on the probability of obtaining significant results.
I	Assessment of interventions by blinded evaluators.
J	Fidelity of the model clearly described.
K	Participants analysed for at least 6 months.
L	Non-measurement of data lower than 20% (studies < 6 months) or 30% (studies > 6 months).
M	Summary of the results of each group and effect size.

Results

The final search yielded a total of 783 results, of which 419 belonged to WOS and 364 to ERIC. After eliminating duplicates (n=62), and considering the four papers added, 725 investigations were reviewed (by reading the title and abstract). Of these, 632 were discarded as it was clear that they did not meet one or more inclusion criteria. The remaining 93 articles were downloaded and reviewed more thoroughly. In the end, a total of seven articles met all the inclusion criteria outlined above. The remaining 86 articles were excluded because they did not meet one or more of the inclusion criteria. The flow diagram (see Supplementary Material II, available at <https://osf.io/438za/files/osfstorage>) schematically shows this process.

Risk of bias in the studies

Five studies passed the risk-of-bias assessment as none of them obtained more than three negative ratings on the items, which means that all of them had a low risk of bias as they scored over 70% (Downs & Black, 1998). However, two of them (Madou et al., 2023; Yan et al., 2024) did not pass this assessment as they received seven and nine negative ratings, respectively. However, we decided to include these



investigations in the review because we considered that both could provide some interesting information as they met the inclusion criteria. Risk of bias in these articles must be considered when interpreting their results. A description of risk-of-bias assessment can be found in Supplementary Material I.

Study characteristics

The characteristics of the studies are shown in detail in Supplementary Material III (available at <https://osf.io/438za/files/osfstorage>). The studies were conducted in 70 primary schools and 22 secondary schools, including as participants 114 teachers (ranging from 7 to 23) and 3,584 students (ranging from 293 to 1,158). While the longest study lasted 10 months, the shortest lasted four weeks ($\bar{x}$ = 5.93 months approximately). Regarding the results obtained, the papers with experimental designs (6/7 studies) achieved positive results when comparing teachers in the control groups (CG) with the experimental groups (EG) (see Supplementary Material III for more details). The studies that included students as participants (5/7) also showed benefits for them in terms of learning, psychological satisfaction, physical activity levels, basic life support performance, engagement, or academic achievement, above all in the EG.

Observational instruments used for the analysis of teaching competencies

Object of assessment and instruments validation method

The analysis of the observational instruments used to assess teaching competencies was an important part of our research purpose. The study by Kyriakides and Tsangaridou (2008) used the ALT-PE instrument, which measured the teachers' time management during the lessons, as well as the quality of the proposed tasks. The observer had to make two decisions: one on the context created by the teacher and the other on the level of student involvement. The record on the context was coded according to whether it was general content (transitions, class management, breaks, etc.), knowledge of the subject content (information concerning techniques, strategies, rules, etc.) or motor content in the subject (students' motor engagement time). ALT-PE was previously developed by Siedentop et al. (1982) and the alpha reliability coefficient was 0.88.

Cheon et al. (2012) used a rating sheet for direct observation. Four raters recorded the teachers' classroom instruction in terms of autonomy support and controlling teaching. The instrument had four items, presented in a bipolar format. The items analysed the encouragement of motivation, the language used, the rational explanations given and the teachers' stance towards negative emotions. The instrument was previously validated (Reeve et al., 2004): alpha coefficients during each classroom observation were high (0.81, 0.90, 0.82), so internal consistencies were high; median interrater reliability was 0.80 for autonomy-supportive instructional behaviours and 0.76 for engagement behaviours; and the observers' ratings correlated with the students' self-reports (validity).

The article by Escriva-Boulley et al. (2018) describes the use of an observational instrument made up of six dimensions: autonomy support, control, structure, chaos, involvement, and hostility. The raters rated each dimension considering the number of times each strategy appeared and the quality of the strategy. The validation process was not specified. The study only mentions that the instrument was based on Self-Determination Theory and on Skinner and Edge's (2002) work.

McKenzie et al. (1993) used the SOFIT instrument to obtain information on the quantity and quality of instruction in PE lessons. Observers evaluated curriculum implementation in terms of student activity level, curricular context of the lesson and teacher behaviour. To assess the curricular context, the observers decided whether the class was in a general content moment (such as classroom management) or a subject knowledge moment. If a PE subject knowledge moment occurred, the evaluators had to decide whether it was content knowledge or motor (physical activity) content. If the latter occurred, a decision had to be made between fitness, skill practice, or games. Teacher behaviour was coded into six categories: promotes fitness, demonstrates fitness, generally instructs, manages the class, observes, and is absent. The SOFIT instrument was validated two years earlier and inter-observer reliability was 0.90 (McKenzie et al., 1991).

In the Miller et al. (2017) research, Quality Teaching Lesson Observation Scales were used. These scales assessed teacher behaviour with respect to three dimensions: intellectual quality, quality of the learning



environment, and significance. This instrument made it possible to focus on both the teacher and the students, to code measures that focused on pedagogy and content, and to obtain an overview of teacher behaviour during the teaching process. The article does not directly mention the validation process, but only explains that the instrument belongs to the New South Wales Department of Education and Training.

Yan et al. (2024) used the MASTER Coach Observation Tool to evaluate teaching behaviours in terms of practice form (% of activity time) and teacher feedback (frequency of times). They also analysed the active learning time. They explained that this tool is a modified version of the Coach Analysis Intervention System (Cushion et al., 2012), but the validation process is not described.

Madou et al. (2023) coded teachers' use of time and teachers' interactions. The use of time was collected by categorizing the length of time all the students were engaged: general context, subject matter knowledge content and subject matter motor content. Teacher-student interactions were coded in two categories: interactions with the whole class and interactions with individual students. The validation process is not mentioned.

Observation Method

In three of the studies, data recording was performed using intervals. In one of them, observations were made at five-second intervals: five seconds for observation and five seconds for recording (Kyriakides & Tsangaridou, 2008). In another, coding was performed every 20 seconds (McKenzie et al., 1993). Escrive-Boulley et al. (2018) divided the lessons into four parts and the observers had to take one recording for each category in each part. Madou et al. (2023) and Yan et al. (2024) used event recording technique and recorded the use of time during lessons. On the other hand, five of the investigations reflected recording independence in their raters, i.e., at least two observers simultaneously assessed the same situation and made separate recordings (Cheon et al., 2012; Escrive-Boulley et al., 2018; McKenzie et al., 1993; Miller et al., 2017; Yan et al., 2024). Madou et al. (2023) described this only for students' observation. In the Cheon et al. (2012) and Escrive-Boulley et al. (2018) investigations, the observers were blinded to treatment conditions, while in Madou et al. (2023) this only occurred for students' observation. In Miller et al. (2017) this was also the case, apart from the principal investigator, who conducted the assessment of intervention fidelity during observations. Only two of the articles explained that the assessment of the lessons was unannounced, so teachers did not know in advance when they were going to be observed (Cheon et al., 2012; McKenzie et al., 1993). The lessons that were videotaped and analysed later were those of Escrive-Boulley et al., (2018), Madou et al. (2023) and Yan et al. (2024).

Observer Preparation

Observer preparation was conducted in all seven investigations, although two of them did not describe the process (Madou et al., 2023; Yan et al., 2024). The most recurrent method of preparation was the use of videos. Observers had to obtain 85% (McKenzie et al., 1993) or 95% (Miller et al., 2017) agreement with pre-coded standards. The raters in the Kyriakides and Tsangaridou (2008) study practiced coding techniques also through video analysis, and in Escrive-Boulley et al. (2018) they had to score two pilot videos using the instrument and compare scores with the principal investigator. Other preparation methods were also used. Observers studied categories defined for behaviour, coded procedures, practiced direct observation (Kyriakides & Tsangaridou, 2008), viewed PowerPoint presentations, conducted small group seminars, participated in collaborative and independent coding lessons (Escriva-Boulley et al., 2018), and conducted discussions, role-plays, and field observations (McKenzie et al., 1993). The article by Cheon et al. (2012) does not give information regarding the preparation of the evaluators, although it does specify that there was preparation.

Inter-observer reliability

Inter-observer reliability was high in all the studies. In Miller et al. (2017), the observers achieved 90% inter-observer agreement and negotiated the items where there was no consensus. When agreement was not possible, the lower score was used. McKenzie et al. (1993) planned to retrain observers when inter-observer agreement was less than 80%, but observers achieved reliabilities of 92.9%. Inter-observer agreement in Escrive-Boulley et al. (2018) was 92%. The inter-rater reliability coefficient p_2 was 0.84 in Kyriakides and Tsangaridou (2008). In Cheon et al. (2012) the ratings of the two observers



were highly correlated: 85.5%. Due to this high reliability, the researchers averaged the two ratings into a single score for each instructional behaviour. In Yan et al. (2024) the two observers averaged 93% inter-rater reliability, while Madou et al. (2023) reported values between .81 and 1 in Cohen's kappa.

Teaching competencies in the studies

Teaching competencies were also a focus of the research objective. The seven studies selected in this systematic review show evidence, either directly or indirectly, of the competencies that teachers have when teaching PE classes. In total, 44 competencies (some of them repeated) were analysed among all the studies, ranging from two competencies to 18 throughout the studies. Supplementary Material IV details all the analysed competencies (available at <https://osf.io/438za/files/osfstorage>).

The competence to provide feedback to students is the most common, appearing in four studies. In Escriva-Boulley et al. (2018) precisely the opposite competence is assessed, i.e., "chaos" (no feedback). In this case, no differences in this regard were observed between the two groups. In contrast, in Miller et al. (2017), teachers in the EG increased feedback regarding their students and their learning process. Madou et al. (2023) reported more interactions for the specialized content knowledge plus common content knowledge (SCK+CCK) group on average, compared to the CCK group. In Yan et al. (2024), MASTER teachers provided students with significantly more effective feedback than CG teachers.

Lesson management appeared in three of the investigations. In the study conducted by McKenzie et al. (1993) no statistically significant differences were observed between the CG and intervention groups with respect to lesson management. Kyriakides and Tsangaridou (2008) pointed out that general content (which includes classroom management) had significant effects on student achievement. In Madou et al. (2023) management was focused on time use, with the SCK+CCK group spending 70% of the time on subject matter motor content compared to 56% in the CCK group. Another competence found in three of the studies is that of involving the students in tasks. The intervention carried out in one of these studies (Miller et al., 2017) had positive effects, since the teachers who participated in the intervention improved the quality of the learning environment in their lessons, with greater student involvement in classroom activities. The study by Kyriakides and Tsangaridou (2008) analysed PE quality from a multilevel perspective, and it was at the classroom level where this teaching competence could be found. Yan et al. (2024) described no significant differences between groups in involving students in tasks, measured in terms of active learning time.

There are some competencies that appear in two articles. The competence to increase physical activity time during the lesson is in Escriva-Boulley et al. (2018) and McKenzie et al. (1993), in which the physical activity levels of the students belonging to the EG were higher than those in the CG. The ability to increase student motivation also appears in two studies, in which teachers were able to increase their students' motivation after receiving preparation on a game-based model (Miller et al., 2017) and on autonomy-building (Cheon et al., 2012). The competence to not have a controlling style or language is found in Cheon et al., (2012) and Escriva-Boulley et al. (2018). The former showed that the EG teachers used more informative language, while in the latter no significant differences were found in terms of the degree of control by teachers between the two groups. The competence to provide good explanations attained positive results, since teachers in the intervention groups provided more rational and effective explanations in the instruction of key concepts (Cheon et al., 2012; Miller et al., 2017). The research by Cheon et al., (2012) and Escriva-Boulley et al. (2018) analysed the competence to support learner autonomy, where the preparation received by teachers proved to be effective, as both articles highlight high levels of autonomy support in the EG.

Discussion

From the analysis of the results presented in this review it can be deduced that the problem of quality of teaching in PE, studied from an observation-based perspective through randomized studies, has hardly been addressed. The specific aim and criteria of this study have meant that some observational studies on teaching competencies in PE have been excluded. In addition, the diversity of the theoretical frameworks that support the studies and the objectives that are proposed, as well as the rest of the characteristics analysed, make it difficult to draw conclusions with clear evidence concerning teaching competencies. However, we consider that both, the demanding inclusion criteria and the large amount



of information extracted from the studies analysed, provide us with information of sufficient quality to advance in the design and approach of future studies that allow scientific development to improve the profession of PE teachers.

Observational instruments and procedure development

Studies on the effectiveness of PE teaching that have employed external evaluation have mainly focused on analysing the students' participation time as a measure of their learning effectiveness (Metzler, 1989; Silverman & Skonie, 1997) without considering the teacher's actions in the spotlight, although there are some validated observational instruments that focus on different aspects of PE (Moon et al., 2022; Roberts & Fairclough, 2012). However, none have been able to compile all the core competencies necessary to define quality teaching.

A more in-depth analysis of our review indicates that observational tools have been used for the analysis of teacher and student behaviour in classroom settings. Some studies have provided important information on the external evaluation of teaching quality in the application of teaching models based on Self-Determination Theory (Cheon et al., 2012; Escriva-Boulley et al., 2018) and in preparation programs based on the change in the teaching of PE from models based on motor skills to models based on play (Miller et al., 2017; Yan et al., 2024). Nevertheless, their application in the study of the quality of PE teaching provides limited frameworks for action, offering an incomplete picture of the current state of the problem as well as not being able to analyse its behaviour in different areas of teacher PD, such as initial or continuing teacher preparation.

The second use of observational instruments was oriented towards the analysis of the behaviours of the PE teachers in the classroom from a broad perspective, focusing on the study of teaching quality from a generic and open conception. The study by McKenzie et al. (1993) used the SOFIT instrument, which has been applied in other research (Powell et al., 2016), and enabled the assessment of both the quantity and quality of instructions given in PE classes by teachers. Although this instrument originally had a rather open conception, SOFIT has evolved to SOFIT+ (and other variations) in order to yield detailed information about teacher practices related to children's moderate-to-vigorous physical activity (Weaver et al., 2016). The scales for observing the quality teaching of lessons (Miller et al., 2017) is designed to assess a teacher's behaviour via specialized categories. This tool also adopts a wide perspective, but it picks up more detailed information about teacher practices (compared to the original SOFIT) concerning three dimensions: intellectual quality of teacher behaviour, quality of the learning environment and meaningful learning. The ALT-PE (Kyriakides & Tsangaridou, 2008) measured the teachers' time management during the lessons and the quality of the proposed tasks. As with SOFIT, ALT-PE and its variations have led to many other instruments being developed (Ward et al., 2023).

From this exposition we can see that all the selected studies propose the analysis of the teacher's behaviour in the classroom setting. However, none of them attempts to analyse teaching quality in a comprehensive way, focusing on the essential competencies that teachers should possess. For this reason, we propose the development of a comprehensive observational instrument. This instrument should have a broad scope and not be limited to assessing specific models. Its purpose must be to determine if PE teachers deliver quality teaching. The instrument's development should examine the teachers' practices in the distinct phases present within PE lessons: the teacher action phase and the student action phase. During the former, the teacher assumes the lead role in explaining activities or content, whereas the student action phase entails the moment when the students actively participate in activities. Understanding how teachers execute instruction, their behaviour during student activities, and the curriculum's interrelation forms a strong framework for assessing teaching quality. It is also important to consider factors such as motivation, inclusion, coeducation, the use of innovative technologies, and evaluation. Additionally, to provide more context, it would be beneficial to gather data on the motor activities created and the students' organization. Such an instrument would be valuable and would serve as a foundation for a global assessment of the quality PE teaching, and results extracted from this review help to fulfil this need.

Regarding the observation procedure, the studies analysed describe the use of recording techniques using time intervals, event recording, simultaneous evaluation by at least two observers, who make independent assessments, blinding of the observers with respect to the treatment conditions, unannounced evaluation of the lessons, and video recording of the intervention, which are practices



commonly used by researchers in observational methods (Moon et al., 2022; Roberts & Fairclough, 2011). Additionally, the majority of observations conducted in the selected papers met some scientific guarantees, such as the validity of the observation instrument, the reliability and control of sources of error and the degree of agreement between observers (García Sánchez et al., 2010). Nevertheless, although not all of the studies used video recording, technological progress in recent years has made possible the development of multiple applications that allow all perceptible behaviours to be recorded, and to display the recordings as frequently as necessary.

With regard to evaluator preparation, the competence of the observer is an element to be considered, since the three components of the observation (perception, interpretation and prior knowledge) can cause different biases that must be controlled (García Sánchez et al., 2010). In the case of this review, in all the investigations, observer preparation was carried out to accurately record what occurred during the lessons. In many studies, observer preparation was similar, with theoretical preparation on coding and working by categories and viewing videos, and paying attention to intra- and inter-observer reliability (Curtner-Smith & Todorovich, 2002; Escartí et al., 2018; Roberts & Fairclough, 2011). However, preparing evaluators is a challenging process. It is difficult to find external evaluators who are willing to commit to an extensive preparation period. Additionally, the preparation may not yield the expected results in terms of intra- and inter-observer reliability. Therefore, we advocate for standardizing training protocols to ensure validity and reliability in observer preparation.

It is remarkable that, despite the disparity of instruments used, the results show that all the interventions for the improvement of the quality of teaching provide positive results, which reinforces the idea of the importance of using valid, viable and reliable instruments, and that appropriate preparation of the observers is required.

Teaching competencies

The selected studies analyse a total of 31 different teaching competencies. The use of these competencies is identified, in this research, as an important part of quality teaching. The competence that appears most often in the articles is providing feedback to students (4/7). The ability to provide feedback to students during the activities is essential because it is effective and students can use this information to improve their learning when necessary (Ensign et al., 2017), so PE teachers must be able to provide feedback to their students, which is an indicator of great PCK (Backman & Barker, 2020).

Lesson management appears in three studies. In a context such as PE, it is useful to know how the time is managed during the lesson, since the particularities of the subject (spaces, materials, etc.) can lead to the loss of valuable time, and effective teachers must be able to anticipate events, make quick transitions and, ultimately, maximize motor engagement time (Rink & Hall, 2008). This competence is one of the most important ones (Rink, 2013) and acquires special importance for most of the participants in the study conducted by Ensign et al. (2018), who consider it to be a crucial characteristic within the construct of teacher effectiveness.

The competence to involve students in tasks is also present in other studies, and it is therefore included in some instruments to measure teaching quality, such as that of Rink (2013). This student engagement is recognized as fundamental in the learning process; it is necessary to reflect, analyse, debate and take decisions about this process, which will contribute to meaningful learning (Ausubel, 2002). To involve students in tasks, it is essential to foster their motivation, which is another observed competence. Similar to our selected studies, Perlman and Karp (2007) also obtained positive results, as the participating teacher trainees also managed to create a good motivational environment in their classrooms.

The competence to increase the time of physical activity during the lessons was analysed by other researchers like Powell et al. (2016), who demonstrated that the students who received the intervention showed higher levels in this regard, as occurred in our selected studies. However, it is essential to consider that physical activity time is not a guarantee of student learning, so teachers must ask themselves what to do to create situations that allow the students to learn.

The competence to not have a controlling style or language showed positive results in Cheon et al., (2012) (less controlling style in the EG) but no significant differences were reported in Escrivá-Boulley et al. (2018). The case of teachers in initial preparation is a little different, since they know that they

should not exercise a controlling style in their classes, but if they give the students freedom they feel that they should be managing the situation more, meaning that more autonomous strategies are not implemented in their lessons (Perlman & Karp, 2007). Related to this, the research by Cheon et al., (2012) and Escriva-Boulley et al. (2018) analysed the competence to support learner autonomy, obtaining higher levels of autonomy support in the EG. However, the results obtained by Ko et al. (2006) were different, since most of the teachers in their study were not able to give more responsibility to their students during class, despite having programmed this. These results might be explained for the same reason pointed out in Perlman and Karp (2007), where pre-service teachers felt that they should have been managing the class, maintaining control over the students.

The competence to provide good explanations achieved favourable results in the selected studies intervention groups, which are congruent with those obtained by García-Fariña et al. (2018), whose participants improved the use of their discursive strategies after receiving preparation related to this. One of the keys to know if explanations have been adequate is to observe the behaviour of the students, to see if they are involved in the activity (Rink, 2013). The novice teachers in the research by Ensign et al. (2017) were also able to engage their students in the tasks, providing clear introductions for the activities.

As already noted, there are other teaching competencies which also appear in some of the seven studies that have been included in this review, but that have not been collected in more than one of the studies; however, this does not mean that they are less important. Despite the total number of teaching competencies collected (31) (see Supplementary Material IV), we can say that focusing exclusively on these skills is not enough to fully evaluate effective teaching in PE. To achieve this goal, we should also consider competencies such as respectful communication, the fostering of divergent thinking, coeducation implementation, appropriateness of new technology use or final reflection development, since they are considered essential in teaching quality. Furthermore, we recommend future researchers into teaching quality to evaluate not only the competencies observed in PE lessons directly, but also the fundamental skills related to the scheduling of lessons or to PD and commitment. We are aware that we still have a long way to go before we can define a set of fundamental teaching competencies for the development of teaching in PE, but we consider this review could contribute to define core in-class teaching competencies which teachers must possess.

Theoretical background for the instruments used

The studies selected in this systematic review use seven different instruments for the assessment of teaching competencies. Except for McKenzie et al., (1993) and Madou et al. (2023), the instruments used responded to a particular theoretical framework, so that the competencies analysed corresponded to these. Kyriakides and Tsangaridou (2008) based their study on Creemers' (1994) model, which is an extension of Carroll's (1963) model. This model assesses educational quality from a multilevel perspective (context, school, and classroom) and divides learning time into three variables: time allowed to learn, time in which students are willing to actively engage in learning, and time required to learn under optimal instructional conditions. Therefore, the instrument assessed competencies for teacher management of the lesson and for involving students in the activities. The research by Cheon et al. (2012) and Escriva-Boulley et al. (2018) based their assessment instruments on Self-Determination Theory (Deci & Ryan, 2000). According to this theory, human motivation requires the consideration of three innate psychological needs: competence, autonomy, and relatedness. Consequently, the instruments used analysed competencies such as supporting students' autonomy, fostering their motivation or the absence of a controlling style. Miller et al. (2017) and Yan et al. (2024) principally based their studies on the game-centred approach, which understands play as a medium for learning (Kirk & MacPhail, 2002), and assessed competencies such as approaching knowledge from different perspectives, creating a supportive social climate, developing decision making activities, and using stories.

The fact that the various tools used are based on specific theoretical models has limited their scope. Therefore, our proposal to develop a global observational instrument for the comprehensive evaluation of teaching in PE should not necessarily respond to a specific model but should be able to be used by different models that have the classroom as an object of study. What we consider important to assume is that teachers are one of the central elements of the classroom environment (along with the students and the content) (Dunkin & Biddle, 1974), and the interdependence of these elements characterizes the



classroom as an ecological system in which each element is mutually reinforced and interconnected. Consequently, a change in one element can elicit changes in all the others (Doyle, 1986). From this ecological paradigm of classroom vision, the teacher's behaviours in the classroom are a basic element of the teaching quality system. They are not considered as mere product variables evaluated based on the results of their students. Rather, they are process variables that have intrinsic value and that, in interaction with the other elements of the classroom system, condition teaching effectiveness. The new tool for a global assessment of PE teaching should assume that teacher behaviour plays a central role in teaching quality, and this instrument could be applied from various theoretical frameworks that have the classroom as their object of study.

As for the limitations of this review, we can point out that only articles belonging to the WOS and ERIC databases were analysed, so that some articles that may have passed all the inclusion criteria could have been left out of our selection. In addition, the small number of selected studies means that some important competencies (described earlier) did not appear, so it was not possible to analyse them. Related to this, the lack of studies with more homogeneous objectives and methodologies is probably the most important factor that has limited the possibility of drawing more solid conclusions to advance the study of the quality of PE teachers teaching. Finally, we would like to include our perspective as a limitation. As researchers with experience and a professional interest in this area, we believe that our perspective concerning how we have handled the information and drawn conclusions from it, especially when analysing and reflecting on the studies and their theoretical frameworks, could be a potential source of bias.

It is clear from the results of this study that there is a need for the scientific community to focus its interest on agreeing on a standard model of evaluation; one that will allow us to gather global information on the quality of teaching, in order to know where we are and thus be able to make decisions about what is needed in the PD of teachers in a changing and complex society. However, although there are high-quality studies that use teacher observation in PE, few of them use a randomised sample, which is an important and widely recognised quality criterion in the scientific field. Therefore, more high-quality studies using randomised samples are needed to assess teaching quality at different stages and in specific contexts, in both initial and ongoing teacher training.

Conclusions

To the best of our knowledge, no other reviews addressed the issue of PE teaching competencies using observational instruments. This study provides information about seven studies from which future research could stem to develop observational tools to assess PE teachers' classroom practices. This review shows that quality studies on PE teaching competencies using observational instruments are scarce.

Regarding the observational instruments used for the assessment of teaching competencies, we conclude that there are a wide variety of valid tools that fit the purposes of each research, but that generally do not address the problem of teaching competencies from a more global and holistic perspective that allows us to study the essential elements of teaching PE. Therefore, we emphasize the need to develop an observational instrument that includes a global perspective to comprehensively assess teaching competencies in PE. This would require a documentary study to compile a greater number of observation tools, and consequently a greater variety of teaching competencies.

The most common competence in the selected studies was providing feedback to students. Other common competencies in the selected studies are being able to manage the lesson, to involve students in tasks, to increase physical activity, to increase student motivation, not to have a controlling style or language, to provide good explanations, and to support learner autonomy. However, even though a comprehensive collection of teaching competencies has been gathered, it is insufficient to completely assess the effectiveness of PE teaching simply by focusing on these skills. To attain this objective, it is important to contemplate competencies such as using respectful communication, fostering divergent thinking, implementing coeducation, the appropriateness of using new technologies, or developing final reflection.



Finally, all the studies that evaluated PD programs for changes in teachers' teaching practice, and those that evaluated student outcomes, attained positive results following interventions, which reinforces the idea that high-quality PD programs are needed and beneficial.

Financing

This research was funded by UV Vice-Rectorate Research Program, Call for Special Actions, grant number UV-INV-AE-3656932.

References

- Asun, S., Chivite, M. T., & Romero, M. R. (2020). Perceptions of professional competences in Physical Education Teacher Education (PETE). *Sustainability*, 12(9): 3812. <https://doi.org/10.3390/su12093812>
- Ausubel, D. (2002). *Adquisición y retención del conocimiento: una perspectiva cognitiva*. Paidós.
- Backman, E., & Barker, D. M. (2020). Re-thinking pedagogical content knowledge for physical education teachers—implications for physical education teacher education. *Physical Education and Sport Pedagogy*, 25(5), 451–463. <https://doi.org/10.1080/17408989.2020.1734554>
- Berckley, J. (2019). What's the Difference Between Teacher Quality and Quality Teaching? In Dr. Jay Berckley. Available at: <https://jayberckley.com/research/teacher-quality-quality-teaching>
- Carroll, J. B. (1963). A model of school learning. *Teacher College Record*, 64(8), 723–733. <https://doi.org/10.1177/016146816306400>
- Cheon, S. H., Reeve, J., & Moon, I. S. (2012). Experimentally based, longitudinally designed, teacher-focused Intervention to help physical education teachers be more autonomy supportive toward their students. *Journal of Sport and Exercise Psychology*, 34(3), 365–396. <https://doi.org/10.1123/jsep.34.3.365>
- Creemers, B. P. M. (1994). *The effective classroom*. Cassell.
- Curtner-Smith, M. D., & Todorovich, J. R. (2002). The Physical Education climate assessment instrument. *Perceptual and Motor Skills*, 95, 652–660. <https://doi.org/10.2466/pms.2002.95.2.652>
- Cushion, C., Harvey, S., Muir, B., & Nelson, L. (2012). Developing the coach analysis and intervention system (CAIS): Establishing validity and reliability of a computerised systematic observation instrument. *Journal of Sports Sciences*, 30(2), 201–216. <https://doi.org/10.1080/02640414.2011.635310>
- De Souza, J., Brasil, V., Kuhn, F., de Barros, T. E., & Ramos, V. (2017). As crenças de graduandos em educação física sobre o ensino dos esportes. *Movimento. Revista de Educação Física Da UFRGS*, 23(1), 133–146. <https://doi.org/10.22456/1982-8918.64032>
- Deci, E. L., & Ryan, R. M. (2000). The “what” and “why” of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry*, 11(4), 227–268. https://doi.org/10.1207/S15327965PLI1104_01
- Downs, S. H., & Black, N. (1998). The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. *Journal of Epidemiology and Community Health*, 52(6), 377–384. <https://doi.org/10.1136/jech.52.6.377>
- Doyle, W. (1986). Classroom organization and management. In M. Wittrock (Ed.), *Handbook of research on teaching* (pp. 392–431). Macmillan.
- Dunkin, M., & Biddle, B. (1974). *The study of teaching*. Holt, Rinehart and Winston.
- Ensign, J., Woods, A. & Kulinna, P. (2017). Teacher Development in First-Year Physical Educators: A Comparison of Effectiveness Among Different Physical Education Teacher Education Backgrounds. *Journal of Teaching in Physical Education*, 36(4), 1–32. <https://doi.org/10.1123/jtpe.2016-0105>
- Ensign, J., Woods, A., Kulinna, P. & McLoughlin, G. (2018). The teaching performance of first-year physical educators. *Physical Education and Sport Pedagogy*, 23(6), 592–608. <https://doi.org/10.1080/17408989.2018.1485140>
- Escartí, A., Llopis-Goig, R., & Wright, P. (2018). Assessing the Implementation Fidelity of a School-based Teaching Personal and Social Responsibility Program in Physical Education and Other Subject



- Areas. *Journal of Teaching in Physical Education*, 37(1), 12–23. <https://doi.org/10.1123/jtpe.2016-0200>
- Escriba-Boulley, G., Tessier, D., Ntoumanis, N., & Sarrazin, P. (2018). Need-supportive professional development in elementary school physical education: Effects of a cluster-randomized control trial on teachers' motivating style and student physical activity. *Sport, Exercise, and Performance Psychology*, 7(2), 218–234. <https://doi.org/10.1037/spy0000119>
- Fitchett, P. G., & Heafner, T. L. (2018). Teacher Quality or Quality Teaching? Eighth Grade Social Studies Teachers' Professional Characteristics and Classroom Instruction as Predictors of U.S. History Achievement. *RMLE Online*, 41(9), 1–17. <https://doi.org/10.1080/19404476.2018.1514826>
- García-Fariña, A., Jiménez-Jiménez, F., & Anguera, M. T. (2018). Observation of Communication by Physical Education Teachers: Detecting Patterns in Verbal Behaviour. *Frontiers in Psychology*, 9(334). <https://doi.org/10.3389/fpsyg.2018.00334>
- García Sánchez, J. N., Pacheco Sanz, D. I., Díez González, M. del C., & García-Martín, E. (2010). La Metodología Observacional Como Desarrollo De Competencias En El Aprendizaje. *International Journal of Developmental and Educational Psychology*, 3(1), 211–217. <https://www.redalyc.org/articulo.oa?id=349832326022>
- Goesling, B. (2019). A Practical Guide to Cluster Randomized Trials in School Health Research. *Journal of School Health*, 89(11), 916–925. <https://doi.org/10.1111/josh.12826>
- Hiebert, J., & Morris, A. K. (2012). Teaching, rather than teachers, as a path toward improving classroom instruction. *Journal of Teacher Education*, 63(2), 92–102. <https://doi.org/10.1177/0022487111428328>
- Kawuryan, S. P., Sayuti, S. A., Aman, & Dwiningrum, S. I. A. (2021). Teachers Quality and Educational Equality Achievements in Indonesia. *International Journal of Instruction*, 14(2), 811–830. <https://doi.org/10.29333/iji.2021.14245a>
- Kennedy, M. M. (2006). From Teacher Quality to Quality Teaching. *Educational Leadership: Journal of the Department of Supervision and Curriculum Development, N.E.A.*, 63(6): 14–19. <https://www.ascd.org/el/articles/from-teacher-quality-to-quality-teaching>
- Kirk, D., & MacPhail, A. (2002). Teaching games for understanding and situated learning: Rethinking the Bunker-Thorp model. *Journal of Teaching in Physical Education*, 21(2), 177–192. <https://doi.org/10.1123/jtpe.21.2.177>
- Ko, B., Wallhead, T., & Ward, P. (2006). Professional development workshops - What do teachers learn and use? *Journal of Teaching in Physical Education*, 25(4), 397–412. <https://doi.org/10.1123/jtpe.25.4.397>
- Kunter, M., Klusmann, U., Baumert, J., Richter, D., Voss, T., & Hachfeld, A. (2013). Professional competence of teachers: Effects on instructional quality and student development. *Journal of Educational Psychology*, 105(3), 805–820. <https://doi.org/10.1037/a0032583>
- Kyriakides, L., Creemers, B., & Panayiotou, A. (2020). Developing and Testing Theories of Educational Effectiveness Addressing the Dynamic Nature of Education. In J. Hall, A. Lindorff, & P. Sammons (Eds.), *International Perspectives in Educational Effectiveness Research* (pp. 33-70). Springer. <https://doi.org/10.1007/978-3-030-44810-3>
- Kyriakides, L., & Tsangaridou, N. (2008). Towards the development of generic and differentiated models of educational effectiveness: a study on school and teacher effectiveness in physical education. *British Educational Research Journal*, 34(6), 807–838. <https://doi.org/10.1080/01411920802041467>
- Lander, N., Eather, N., Morgan, P. J., Salmon, J., & Barnett, L. M. (2016). Characteristics of Teacher Training in School-Based Physical Education Interventions to Improve Fundamental Movement Skills and/or Physical Activity: A Systematic Review. *Sports Medicine*, 47(1), 135–161. <https://doi.org/10.1007/s40279-016-0561-6>
- Lee, O., Choi, E., Griffiths, M., Goodyear, V., Armour, K., Son, H., & Jung, H. (2019). Landscape of secondary physical education teachers' professional development in South Korea. *Sport, Education and Society*, 24(6), 597–610. <https://doi.org/10.1080/13573322.2019.1612348>
- Liang, J. (2015). Live video classroom observation: an effective approach to reducing reactivity in collecting observational information for teacher professional development. *Journal of Education for Teaching*, 41(3), 235–253. <https://doi.org/10.1080/02607476.2015.1045314>



- Madou, T., Depaepe, F., Ward, P., & Iserbyt, P. (2023). The role of specialised content knowledge in teaching basic life support. *Health Education Journal*, 82(5), 555–568. <https://doi.org/10.1177/00178969231174685>
- Martínez-Angulo, C., Deneb, E., Pindal, C., Gallardo-Fuentes, F., Carter-Thuillier, B. & Peña-Troncoso, S. (2023). Percepción de autosuficiencia hacia la inclusión en el futuro profesorado de Educación Física: un estudio en contexto chileno. *Retos*, 48, 919–926. <https://doi.org/10.47197/retos.v48.97413>
- Maulana, R., Helms-Lorenz, M., & Klassen, R. M. (2023). *Effective Teaching Around the World: Theoretical, Empirical, Methodological and Practical Insights*. Springer. <https://doi.org/10.1007/978-3-031-31678-4>
- McKenzie, T. L. (2010). 2009 C. H. McCloy lecture seeing is believing: Observing physical activity and its contexts. *Research Quarterly for Exercise and Sport*, 81(2), 113–122. <https://doi.org/10.1080/02701367.2010.10599656>
- McKenzie, T. L., Sallis, J. F., Faucette, N., Roby, J. J., & Kolody, B. (1993). Effects of a curriculum and inservice program on the quantity and quality of elementary physical education classes. *Research Quarterly for Exercise and Sport*, 64(2), 178–187. <https://doi.org/10.1080/02701367.1993.10608795>
- McKenzie, T. L., Sallis, J. F., & Nader, P. R. (1991). SOFIT: System for Observing Fitness Instruction Time. *Journal of Teaching in Physical Education*, 11(2), 195–205. <https://doi.org/10.1123/jtpe.11.2.195>
- Metzler, M. (1989). A Review of Research on Time in Sport Pedagogy. *Journal of Teaching in Physical Education*, 8(2), 87–103. <https://doi.org/10.1123/jtpe.8.2.87>
- Miller, A., Eather, N., Gray, S., Sproule, J., Williams, C., Gore, J., & Lubans, D. (2017). Can continuing professional development utilizing a game-centred approach improve the quality of physical education teaching delivered by generalist primary school teachers? *European Physical Education Review*, 23(2), 171–195. <https://doi.org/10.1177/1356336X16642716>
- Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & Group, T. P. (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *Plos Medicine*, 6(7), 1–6. <https://doi.org/10.1371/journal.pmed.1000097>
- Moon, J., Webster, C. A., Herring, J., & Egan, C. A. (2022). Relationships Between Systematically Observed Movement Integration and Classroom Management in Elementary Schools. *Journal of Positive Behavior Interventions*, 24(2), 122–132. <https://doi.org/10.1177/1098300720947034>
- Perlman, D. J., & Karp, G. G. (2007). Using TARGET to enhance preservice classroom teachers' ability to identify and develop a motivational climate in physical education. *Physical Educator*, 64(2), 102–113. <https://ro.uow.edu.au/cgi/viewcontent.cgi?article=1634&context=edupapers>
- Postholm, M. B. (2012). Teachers' professional development: A theoretical review. *Educational Research*, 54(4), 405–429. <https://doi.org/10.1080/00131881.2012.734725>
- Powell, E., Woodfield, L. A., & Nevill, A. M. (2016). Increasing physical activity levels in primary school physical education: The SHARP Principles Model. *Preventive Medicine Reports*, 3, 7–13. <https://doi.org/10.1016/j.pmedr.2015.11.007>
- Reeve, J., Jang, H., Carrell, D., Geon, S., & Barch, J. (2004). Enhancing students' engagement by increasing teachers' autonomy support. *Motivation and Emotion*, 28, 147–169. <https://doi.org/10.1023/B:MOEM.0000032312.95499.6f>
- Rink, J. E. (2013). Measuring teacher effectiveness in physical education. *Research Quarterly for Exercise and Sport*, 84(4), 407–418. <https://doi.org/10.1080/02701367.2013.844018>
- Rink, J. E., & Hall, T. J. (2008). Research on effective teaching in elementary school physical education. *The Elementary School Journal*, 108(3), 207–218. <https://doi.org/10.1086/529103>
- Roberts, S., & Fairclough, S. (2011). Observational analysis of student activity modes, lesson context and teacher interactions during games classes in high school (11-16 years) physical education. *European Physical Education Review*, 17(2), 255–268. <https://doi.org/10.1177/1356336X11420222>
- Roberts, S., & Fairclough, S. (2012). A five-stage process for the development and validation of a systematic observation instrument: The system for observing the teaching of games in physical education (SOTG-PE). *European Physical Education Review*, 18(1), 97–113. <https://doi.org/10.1177/1356336X11430653>



- Setiawan, I., Kurniawan, W. R. & Wijayanti, D. G. S. (2024). Implementación de programas de desarrollo profesional docente en línea: perspectivas de los docentes de educación física. *Retos*, 57, 445–454. <https://doi.org/10.47197/retos.v57.105406>
- Siedentop, D., Tousignant, M., & Parker, M. (1982). *Academic learning time -physical education: coding manual*. Ohio State University.
- Silverman, S., & Skonie, R. (1997). Research on Teaching in Physical Education: An Analysis of Published Research. *Journal of Teaching in Physical Education*, 16(3), 300–311. <https://doi.org/10.1123/jtpe.16.3.300>
- Skinner, E., & Edge, K. (2002). Parenting, motivation, and the development of children's coping. In L. Crockett (Ed.), *Agency, Motivation, and The Life Course* (pp. 77–143). University of Nebraska Press.
- UNESCO (2015). *Quality Physical Education (QPE): guidelines for policy makers*. UNESCO Publishing. <https://unesdoc.unesco.org/ark:/48223/pf0000231101>
- Van Sluijs, E. M. F., McMinn, A. M., & Griffin, S. J. (2007). Effectiveness of interventions to promote physical activity in children and adolescents: Systematic review of controlled trials. *British Journal of Sports Medicine*, 335, 1–13. <https://doi.org/10.1136/bmj.39320.843947.BE>
- Vasconcellos, D., Parker, P. D., Hilland, T., Cinelli, R., Owen, K. B., Kapsal, N., Lee, J., Antczak, D., Ntoumanis, N., Ryan, R. M., & Lonsdale, C. (2020). Self-Determination Theory Applied to Physical Education: A Systematic Review and Meta-Analysis. *Journal of Educational Psychology*, 112(7), 1444–1469. <https://doi.org/10.1037/edu0000420.supp>
- Vattøy, K. D., & Gamlem, S. M. (2020). Teacher–student interactions and feedback in English as a foreign language classrooms. *Cambridge Journal of Education*, 50(3), 371–389. <https://doi.org/10.1080/0305764X.2019.1707512>
- Villaverde-Caramés, E. J., Fernández-Villarino, M. Á., Toja-Reboredo, B., & González-Valeiro, M. Á. (2021). A literature review of the characteristics that define a good physical education teacher: considerations from teacher training. *Retos*, 41, 471–479. <https://doi.org/10.47197/retos.v0i41.84421>
- Ward, P., Ayvazo, S. & Iserbyt, P. (2023). Daryl Siedentop's Epistemological Lens, and His Influence on the Fields of Physical Education and Teacher Education. *Quest*, 75(1), 5-15. <https://doi.org/10.1080/00336297.2022.2132967>
- Weaver, R. G., Webster, C. A., Erwin, H., Beighle, A., Beets, M. W., Choukroun, H., & Kaysing, N. (2016). Modifying the System for Observing Fitness Instruction Time to Measure Teacher Practices Related to Physical Activity Promotion: SOFIT+. *Measurement in Physical Education and Exercise Science*, 20(2), 121–130. <https://doi.org/10.1080/1091367X.2016.1159208>
- Yan, J., Morgan, P. J., Smith, J. J., Chen, S., Leahy, A. A., & Eather, N. (2024). Pilot randomized controlled trial of a game-based intervention for teaching basketball in Chinese primary school physical education. *Journal of Sports Sciences*, 42(1), 25–37. <https://doi.org/10.1080/02640414.2024.2319457>

Authors' and translators' details:

Jesús Mellado-Berenguer
Manuel Monfort-Pañego
Daniel Sheerin

jemebe@alumni.uv.es
manuel.monfort@uv.es
dbsheerin@hotmail.com

Author
Author
Translator

