Improving the quality of Physical Education: a teaching behaviour-based study protocol using artificial intelligence

Mejorando la calidad de Educación Física: protocolo de estudio basado en conductas docentes mediante inteligencia artificial

Authors

Evelia Franco [1]
Daniel Gutiérrez-Reina [2]
Alba González-Peño [3,*]
Javier Coterón [3]

[1] Universidad Loyola Andalucía (Spain)
[2] Universidad de Sevilla (Spain)
[3] Universidad Politécnica de Madrid (Spain)

Corresponding author:
Alba González Peño
alba.gonzalez.peno@upm.es

Recibido: 26-06-25
Aceptado: 22-08-25

How to cite in APA

Franco, E., Gutiérrez-Reina, D., González-Peño, A., & Coterón, J. (2025). Improving the quality of Physical Education: a teaching behaviour-based study protocol using artificial intelligence. *Retos, 72*, 715-727. https://doi.org/10.47197/retos.v72.116921

Abstract

Introduction: Over the last few decades, the analysis of quality in the physical education (PE) context has received remarkable attention, self-determination theory being one of the most successful theoretical perspectives in explaining teacher-student interactions. The aim is to present an artificial intelligence-based study protocol to automatise the identification of teaching behaviours, enhancing educational research oriented to improving PE quality. Method: Eight different teaching behaviours are classified using natural language processing techniques. A data set is generated containing transcriptions of voice recordings from numerous PE lesson extracts, coded by PE experts. These data are used to train different machine learning algorithms so that teaching behaviours can be automatically identified and labelled. Results: Algorithms tested will be assessed through different metrics such as accuracy, precision, recall, and F1-score for each teaching behaviour to be predicted. The findings are believed to provide a promising tool to improve educational research, which will, in turn, favour the quality of PE teaching behaviours. Discussion: The analysis of teaching behaviours and students' outcomes has traditionally relied on self-reported questionnaires and external observation. While valid, these practices are highly time- and resource-consuming, acting as a barrier to sustaining certain projects aimed at improving educational practices. This study protocol seeks to overcome such limitations.

Keywords

Circumplex approach; motivation; natural language processing; supervised machine learning; teaching behaviours.

Resumen

Introducción: En las últimas décadas, el análisis de la calidad del contexto de la educación física (EF) ha recibido una atención significativa, siendo la teoría de la autodeterminación una de las perspectivas teóricas más exitosas en la explicación de las interacciones entre docentes y estudiantes. Este estudio presenta un protocolo basado en inteligencia artificial para automatizar la identificación de conductas docentes, contribuyendo a la investigación educativa orientada a mejorar la calidad de la EF. Método: Ocho tipos de conductas docentes son clasificados mediante técnicas de procesamiento del lenguaje natural. Se genera un conjunto de datos con transcripciones de grabaciones de voz de cientos de fragmentos de clases de EF, codificados por expertos. Estos datos se emplean para entrenar algoritmos de aprendizaje automático que permiten identificar y etiquetar automáticamente las conductas docentes. Resultados: Los algoritmos evaluados se analizarán mediante distintas métricas, tales como precisión (accuracy), exactitud (precision), sensibilidad (recall) y F1-score para cada comportamiento docente que se pretenda predecir. Los hallazgos pueden constituir una herramienta prometedora para mejorar la calidad de la investigación educativa y, en consecuencia, de la enseñanza en EF. Discusión: Además, el análisis de conductas docentes y resultados de estudiantes se ha basado tradicionalmente en cuestionarios autoadministrados y observación externa. Si bien estos métodos son válidos, su elevado consumo de recursos y tiempo dificulta la sostenibilidad de ciertos proyectos educativos. Este protocolo busca superar estas limitaciones.

Palabras clave

Aprendizaje automático supervisado; comportamientos docentes; motivación; modelo circular; procesamiento del lenguaje natural.

Introduction

The quality of physical education (PE) has received growing attention over the last few years by both the scientific community (Dudley et al., 2022; Leão & Lorente-Catalán, 2024; Reyes-Rodríguez et al., 2025) and worldwide organisations (UNESCO, 2015). One reason for such interest is that children and adolescents' experiences in PE class has been identified as a significant determining factor of their own behavioural patterns in the PE setting, which, in turn, will foster physical activity (PA)-related adaptive outcomes among adolescents (Franco et al., 2021; Tilga et al., 2021). This is particularly relevant given that, despite the significant scientific evidence accumulated for years on the positive influence of regular PA on health-related outcomes (Gao et al., 2018), PA levels among adolescents do not reach the recommended standards (Guthold et al., 2020).

Quality Physical Education

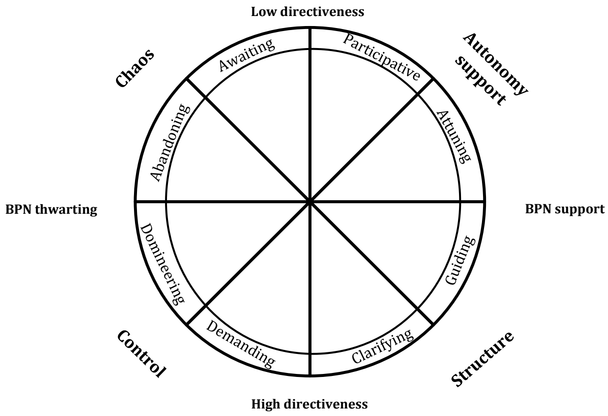
Teachers have generally been identified as key actors in general education quality. In 2007, the McKinsey Report highlighted the importance of the role of teachers, stating that no education system can be better than the quality of its teachers (Barber & Mourshed, 2007). Along these lines, several studies affirm that the agent with the greatest impact on the education system, and therefore on the comprehensive education of students, is the teaching staff (Comisión Europea, 2017; Cudney et al., 2023; Hill et al., 2003; Mourshed et al., 2012). This trend certainly seems to be applicable to the case of PE. Based on qualitative studies, it was suggested that PE had the potential to contribute to the positive development of young people in the aforementioned domains. However, the review also emphasised that many of these benefits may be due to the nature of the interactions between pupils and their teachers. More recently, quantitative meta-analytic findings have highlighted the multi-domain benefits of PE for development (Dudley et al., 2022), and specific instructional techniques have been shown to support learning and developmental outcomes to a stronger or weaker degree.

Self-Determination Theory And The Circumplex Approach As Valid Frameworks To Analyse Interaction In The PE Setting

Self-determination theory (SDT; Ryan & Deci, 2000) is a theoretical framework that has frequently been used in the analysis of the aforementioned interactions (Cádiz et al., 2021; Lamoneda et al., 2024). This theory establishes that the motivation that people can experience fluctuates on a continuum ranging from demotivation to intrinsic motivation. Based on the tenets of this approach, students' motivational regulations in PE are influenced by satisfaction of their basic psychological needs (BPN), namely autonomy (when students can choose the activities in class), competence (when students can achieve the proposed challenges), and relatedness (when students feel confident sharing their thoughts with their peers). What is more, there is overwhelming evidence that teaching behaviours are determinants either to support or thwart each of these needs (Ahmadi et al., 2023).

A fine-grained and innovative approach has recently emerged for a deeper understanding of the interactions between teacher and student through the analysis of teachers' behaviours in the educational context (Aelterman et al., 2019; Escriva-Boulley et al., 2021). This new framework differentiates four teaching styles, characterised by the level of satisfaction or thwarting of BPN and the level of directiveness (see Figure 1). The combination of both axes results in the four teaching styles: autonomy support, structure, control and chaos, which are specified in eight detailed categories depending on the strategies teachers adopt. Specifically, participative and attuning approaches correspond to autonomy support, guiding and clarifying approaches are related to structure, demanding and domineering approaches are associated to control, and abandoning and awaiting approaches correspond to chaos.

Figure 1. Representation of the circumplex model (Aelterman et al., 2019)



Current Approaches In The Analysis Of The PE Setting

Regarding methodology, analysis of the PE setting has been carried out through studies based on self-report measures, mainly using questionnaires completed by students. While these assessment tools are useful when they have adequate reliability and validity given their low cost and ease of administration, there is evidence that the data collected may be biased by the need for social approval conceptualised as social desirability (Larson, 2019). Observational studies have positioned themselves as an interesting alternative that allows us to delve deeper into the analysis of real teaching situations, avoiding this bias. However, the difficulty of collecting data from video recordings in schools and the costly processing of the information (requiring numerous observers to identify and catalogue behaviours) hinder progress in this line of work. Valid as both these approaches might be, these practices are highly time- and resources-consuming as they require researchers to obtain permission to access schools, to administer the instruments, to record sessions, and to watch and code the videos (which must usually be performed by more than one researcher to ensure validity). This is a barrier to the sustainability of certain projects oriented to improve educational practices, since it is difficult to access and analyse the context.

Artificial intelligence (AI) could provide an answer to the aforementioned limitations, and its incorporation into analysis could provide an important boost to research in this area. In recent years we have witnessed a growing boom in the use of AI in PA and sport science research, leading to interesting advances in lines such as sports performance analysis, injury avoidance or identification, and improvement of game systems (e.g., Claudino et al., 2019; Jiang et al., 2016), and even the understanding of motor skills (Mamani-Ramos et al., 2025). However, despite its potential, no studies have incorporated AI in the analysis of teaching behaviours in the PE setting.

The Present Study

Physical education (PE) plays a critical role in promoting students' lifelong engagement in physical activity and healthy lifestyles. Among the various factors influencing PE outcomes, teaching behaviours have received increasing attention for their impact on students' motivation, engagement, and learning experiences. However, traditional methods to assess teaching behaviours—mainly self-reports and direct observations—pose limitations related to subjectivity, feasibility, and scalability. This work aims to present a novel and sustainable AI-based study protocol to address the analysis of PE interactions from the circumplex approach.

Despite the relevance of this approach, there remains a clear empirical gap: to date, no studies have applied AI-based methods to the detection and analysis of teaching behaviours in PE. Existing approaches have relied mainly on self-reports or manual observations, which, while informative, are limited in scope and scalability. Moreover, the technical feasibility of implementing automated solutions in this domain has not yet been tested. Critical challenges—such as speech recognition in noisy environments, the suitability of textual representations for short instructional utterances, and the ability of classification models to capture pragmatic nuances—remain unaddressed. Explicitly acknowledging this gap underscores both the novelty and the exploratory nature of the present protocol.

In this context, the current protocol aims to address three research questions: (1) Can an AI-based system be developed to identify and classify teaching behaviours in PE according to the circumplex model? (2) To what extent do the behaviours identified through this system align with those reported by teachers and perceived by students? (3) What associations can be observed between these teaching behaviours and key student outcomes, including motivational and behavioural indicators? The remainder of the article is structured as follows: the Method section outlines the participants, instruments, and procedure, as well as the data analysis strategies for each aim. Next, the Results section presents the anticipated analyses for each research question. The Discussion considers the potential implications, challenges, and contributions of applying AI in PE contexts. Finally, the Conclusion summarises the key contributions and broader relevance of this study.

Method

Participants

A total of 20 PE teachers in secondary schools in Spain will be involved in the present study. For the first aim of this work, five ordinary PE sessions will be recorded per teacher, resulting in a total of 100 sessions, that will allow us to generate the database to train the algorithm. Given that each PE lesson will be made up of approximately 750 analysis units, we expect a total of 75,000 extracts. To address the second and third aims of this work, the PE students belonging to two groups (n=60) taught by each of the teachers will also be involved in the study, resulting in approximately 1,200 students. The selected groups will belong to levels between the first and fourth years of secondary school, so the age of the students will range between approximately 12 and 15 years.

Instruments

- Teachers' perceptions of teaching styles. A version of the Situations-in-School (SIS) questionnaire (Aelterman et al., 2019) adapted for teachers in the Spanish context in PE (Burgueño et al., 2023) will be used. The scale consists of 12 vignettes of common situations in class. Each situation presents four different reactions corresponding to one of the four broader teaching styles and to one of the eight approaches. Reactions to each situation are provided on a seven-point Likert scale ranging from 1 ("does not describe me at all") to 7 ("describes me extremely well").
- Teaching behaviours. Teaching styles will be assessed by two trained experts following the main statements used in the SIS-PE-coder observation instrument (Van Doren et al., 2023), which includes descriptions of need-supportive and need-thwarting interactions in the PE context that reflect the eight approaches (Aelterman et al., 2019). Two observational experts will be trained to identify one of the eight teaching styles when there is evidence of the behaviour. An additional option will be included to code behaviour that does not match with any of the teaching styles.
- Teachers' contextual variables. Teachers will be asked about age, gender, years of PE teaching experience, if they have ever had students with disabilities, whether they have them now, and whether they feel sufficiently qualified.
- Students' contextual variables. Students will be asked about their age, gender, their PA practice in their leisure time, if they have a disability and what type, and whether they currently have any classmates with disabilities.
- Students' perceptions of teaching styles. A version of the Situations-in-School (SIS) questionnaire (Aelterman et al., 2019) adapted for students the Spanish context in PE (Burgueño et al., 2023) will be used. The scale consists of 12 vignettes of common situations in class. Each situation presents four different reactions corresponding to one of the four broader teaching styles and to one of the eight approaches. Reactions to each situation are provided on a seven-

point Likert scale ranging from 1 ("does not describe my teacher all") to 7 ("describes my teacher extremely well").

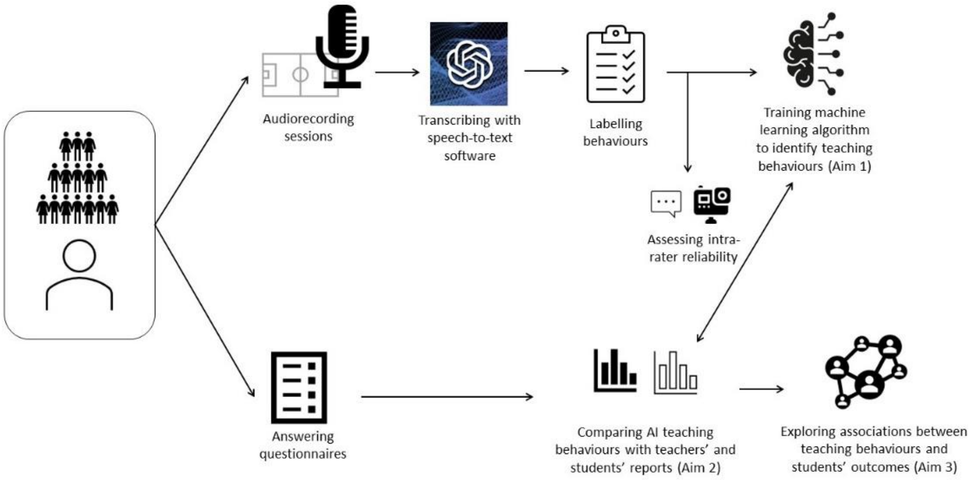
- Satisfaction and frustration of students' BPN. Students' perceptions of satisfaction and frustration of their needs will be assessed by using a Spanish version, adapted to the PE context (Zamarripa et al., 2020), of the scale designed by Chen et al. (2015). The stem used in the questionnaire is "In my PE classes"; this is followed by 24 items grouped in six factors. These six factors, composed of four items, correspond to autonomy, competence and relatedness satisfaction, as well as autonomy, competence and relatedness frustration. Responses will be reported on a five-point scale ranging from 1 ("strongly disagree") to 5 ("strongly agree").
- Behavioural engagement. Students' behavioural engagement will be measured with the Spanish version (Simón-Chico et al., 2023) of the scale adapted from Shen et al. (2012). The stem used in the questionnaire is "In PE classes"; this is followed by five items addressing students' perceptions of their effort, attention, and persistence in PE classes (e.g. "I work as hard as I can"). Responses will be given on a five-point scale ranging from 1 ("strongly disagree") to 5 ("strongly agree").
- Agentic engagement. Agentic engagement will be measured with the Spanish version (Simón-Chico et al., 2023) of the scale developed by Reeve (2013). This instrument is composed of five items that measure the construct of agentic engagement as a single factor (e.g. "During class, I share my preferences and opinions"). Responses will be given on a five-point scale ranging from 1 ("strongly disagree") to 5 ("strongly agree").
- Intention to be physically active. The adapted and translated Spanish version (Moreno et al., 2007) of the Intention to be Physically Active Questionnaire (Hein et al., 2004) will be used. This is composed of five items for measuring the subject's intention to be physically active (e.g. "I am interested in developing my physical fitness"). The items are preceded by the phrase "Regarding your intention to participate in sports...". The questions will be responded to using a Likert scale ranging from 1 ("strongly disagree") to 5 ("strongly agree").

Procedure

Figure 2 summarises the procedure to achieve the aims of the present study. In terms of the data management plan, data will be collected through questionnaires and voice recording. Throughout the entire process, guidelines from the Open Science Framework (https://help.osf.io/) and the American Psychological Association (2002) will be followed. Data will be anonymised and treated with absolute confidentiality during data analysis. To this end, a random code will be generated for each participant so that data collected in different ways can be associated. As a result, each teacher will have a code that will be shared with their students to connect their data. Data will be curated by the research team. If treating data, members of the research team will sign a confidentiality agreement beforehand. Personal data which can lead to the identification of participants will not be shared through open access for confidentiality reasons. In other cases, data will be open access through the European Open Science Cloud (https://eosc-portal.eu/).

In a second phase, data will be collected through questionnaires administered to both teachers and students. Two groups of PE students taught by each teacher will be selected to complete the questionnaires. These will be administered during a PE class, in the presence of a member of the research team, who will explain the purpose of the study, ensure students understand that participation is anonymous and voluntary, and that there are no right or wrong answers. Prior to student participation, written informed consent will be obtained from their parents or legal guardians, in compliance with current data protection regulations and ethical guidelines for research involving minors. Teachers will complete an online version of the questionnaire via the Google Forms platform following the recording sessions.

Figure 2. Procedure summary



Data Analysis To Develop The AI-Based System (Aim 1)

Raw data acquisition

PE teachers and secondary schools, which are included in several databases of educational collaborators, with three different Spanish universities participating in this project, will be contacted. A member of the research team will further contact the PE teachers agreeing to participate in the study to arrange the visit and recording of the session. The voice recording of the teacher will be carried out from the beginning to the end of the PE class, placing a microphone in the lapel of their clothing. Subsequently, this recording will be transferred to OpenAI Whisper software to proceed with the coding of the sessions, which will be carried out by two trained experts. Firstly, to assess the reliability with which teaching behaviours can be assessed through the written text of the teacher's messages, a preliminary concordance analysis will be carried out. To assess intra-rater reliability, one expert, in addition to the written messages, will twice code the video recordings of 20 sessions, two weeks apart. The voice recording will also be processed in speech-to-text software (OpenAI Whisper) and the text will be coded. The resulting coded messages will be used to feed the algorithm. Following previous natural language processing (NLP) approaches, data will be segmented according to utterances, defined as the minimal linguistic entity that can be pragmatically interpreted (Finegan, 2011), and labelled by two experts on the research team according to the teaching behaviours established by the circumplex approach (Aelterman et al., 2019).

Data set creation and organisation

Related to the data set creation, the first challenge in the classification of teaching behaviours lies in the generation of a representative data set from experimental data. As the data set is to be used for training and testing supervised learning algorithms, the different behaviours must be properly labelled. A set of experiments will be carefully designed aiming at an efficient generation, labelling and trimming of the experimental data series in accordance with the different teaching behaviours performed. The procedure will be divided into four different stages: (I) organisation of the teachers and contextual variables; (III) acquisition of the raw data series; (II) detection and segmentation of behaviours; and (IV) target labelling. The raw data acquisition will be conducted through the voice recordings of the five videotaped sessions for each teacher (a total of 50 sessions). Subsequently, the voice data will be transcribed, which will allow the information to be labelled. Transcriptions will be performed using OpenAI Whisper software.

A total of eight different teaching behaviours can be identified in education based on the circumplex approach (Aelterman et al., 2019), namely participative, attuning, guiding, clarifying, demanding, domineering, abandoning and awaiting. These, in turn, can be grouped into four broader categories, namely autonomy-support, structure, control and chaos (Figure 1). Each transcription will be thoroughly analysed by two different researchers with previous training in SDT and the circumplex approach. These two persons will independently identify the smallest units of transcribed messages that can be uniquely framed within one of the teaching behaviours set out in the circumplex approach. In case of discrepancy between the two researchers, they will discuss the section, aiming to reach an agreement. If the discrepancy persists, the decision will be made together with a third member of the research team. Upon completion of codification, the data set will be composed of three different inputs:

a) Smallest units of the transcribed message that can be framed within one specific teaching behaviour (messages containing an unequal number of words).

b) Teaching style identified in the message (coded as 0 = no teaching style associated, 1 = autonomy-support, 2 = structure, 3 = control, 4 = chaos).

c) Teaching behaviour identified in the message (coded as 0 = no teaching behaviour associated, 1 = participative, 2 = attuning, 3 = guiding, 4 = clarifying; 5 = demanding, 6 = domineering, 7 = abandoning, 8 = awaiting).

Once the data have been segmented appropriately, other data transformations are required to work with text data in NLP applications. The following are the data transformations required to work with text.

- Tokenisation. This is the process of breaking down a piece of text, like a sentence or a paragraph, into individual words or 'tokens'. These tokens are the basic building blocks of language. Tokenisation helps computers to understand and process human language by splitting it into manageable units. Tokenisation is a fundamental step in NLP and text analysis tasks, since machine learning and deep learning models take a group of these tokens as the input of the model.

- Stemming. This normalises words into their base or root form. In other words, it helps to predict the parts of speech for each token. It serves to remove redundant information from the text.

- Lemmatisation. This removes inflectional endings and returns the canonical form of a word or lemma. It is like stemming except that the lemma is an actual word.

- Stop words filtering. This consists of removing stop words in the data, since they are frequently used but do not add relevant information.

Data model

Once the data are segmented, structured and labelled according to the teaching behaviours, further data model transformations should be carried out to train machine and/or deep learning algorithms. These algorithms cannot work directly with text data; therefore, numerical transformations should be undertaken—that is, transforming text data into numerical data. The following are data models that are normally used in NLP applications like the one proposed in this work.

- Bag of words. Bag of words (Zhang et al., 2010) uses term frequency (TF) and inverse document frequency (IDF). In this data model, a dictionary or bag of words is created by inserting the most relevant words in the data corpus in a dictionary or 'bag' of terms. The main idea is to use the frequency of appearance of these terms in the document according to target label or teaching behaviours. Thus, if a term or a set of terms appears more frequently in the data segments labelled with a certain teaching behaviour, these patterns are good predictions (good features) of such teaching behaviour. In contrast, if a term appears in all segments regardless of teaching behaviour, it will not be considered a good feature since it does not add useful information to discern between teaching behaviours. Notice that with this data model, the text data are transformed into a numerical vector, obtained as:

$$tf - idf(t,d) = tf(t,d) \times idf(t,d)$$

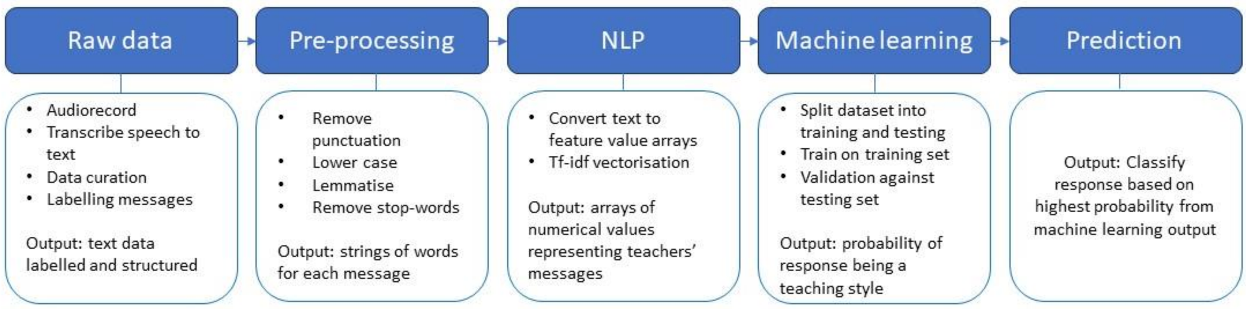
Where *tf(t,d)* is the term frequency of term *t* in the *d* document and *idf(t, d)* is the inverse document frequency of the term *t*. Each term included in the dictionary will be transformed into numerical data using this equation. Furthermore, this data model can be extended by considering n-gram models—that is, using a combination of words as a term in the dictionary.

Word embeddings. Word embedding, or word vector, is an approach to transform data text into a vector representation. It is defined as a numeric vector input that allows words with similar meanings to have the same representation. It can approximate meaning and represent a word in a lower dimensional space. This transformation is crucial since machine learning and deep learning models cannot work with text inputs. Word embeddings are obtained through the training of neural networks with big data text corpus. There are well-known word embeddings such as Word2Vec (Jatnika et al., 2019), GloVe (Pennington et al., 2014), and FastText (Santos et al., 2017), among others. In recent years, transformer techniques based on an attention mechanism have appeared as promising approaches for NLP applications (Vaswani et al., 2017).

Data management for assessment

Once the text data are translated into numerical data by application of the mentioned data models, machine learning and/or deep learning approaches can be used to classify the teaching behaviours. For a proper assessment of the proposed classifier protocol, the available data should be divided into three parts: i) training (60–70%); ii) validation (15%); and iii) testing (15%). Training data are used to train the algorithms. Validation data are used to determine the best configuration or hyper-parameters of the evaluated algorithms. The training data cannot be used to obtain the best hyper-parameters, in order not to bias the results. Finally, test data are used for the final evaluation of the classifier. Therefore, the following standard metrics: accuracy, precision, recall, and F1 of the proposed classifier would be obtained for each category (teaching style), by evaluating the best configuration of the machine learning techniques against the testing data. The entire process of data analysis for Aim 1 is summarised in Figure 3.

Figure 3. The process of analysing teachers' speech



Data Analysis To Compare Teaching Behaviours Identified Through The AI System, Teacher-Reported And Student-Reported Measures (Aim 2)

Firstly, descriptive statistics (means and standard deviations) and correlations among all the study variables will be calculated. A Kolmogorov-Smirnov test will then be performed to verify the normality of the data. According to this normality test, either a parametric analysis of variance (ANOVA) or a Friedman ANOVA for ranked data will be used to analyse differences in teaching style scores as identified through the algorithm and reported by teachers and students. For this aim, different scores will be standardised so that they are comparable. Effect sizes of the comparisons will be estimated using Kendall's W and will be interpreted using Cohen's d interpretation guidelines (Tomczak & Tomczak, 2014). Statistical Package for the Social Sciences (SPSS) version 29 will be used to perform the analyses.

Data Analysis To Explore The Association Between Teaching Behaviours And Students' Outcomes (Aim 3)

As for the third aim of the study, three different stepwise regression analyses will be performed using students' BPN satisfaction and frustration, behavioural engagement, agentic engagement, and intention to be physically active as dependent variables. Teaching styles as identified by the developed algorithm and reported by teachers and students will be treated as independent variables. Statistical Package for the Social Sciences (SPSS) version 29 will be used to perform the analyses.

Discussion

Despite the growing interest in the analysis of teaching behaviours in PE (Cheon et al., 2023), current approaches used in the analysis of PE setting present certain limitations regarding access and analysis of the context, and these could hinder the feasibility and sustainability of certain projects aiming to explore and improve educational practices. This study protocol aims to advance this research line by implementing AI techniques for the identification and classification of such behaviours, with the final purpose of improving the quality of educational practices in the PE context.

The relevance and novelty of the current contribution lie in the fact this is the first attempt to incorporate AI in the detection of teaching behaviours and in the analysis of the interactions occurring in the PE class. So far, teaching behaviours and interaction in the PE class have been explored mainly through self-reported measures (usually questionnaires answered by both teachers and students) and, more recently, through the observation of recordings. These methodological approaches have allowed progress to be made in the knowledge of individual student patterns. More specifically, we propose the use of deep learning techniques to identify and classify teacher behaviour from verbal messages during class. Previous works in other domains such as healthcare, finance or smart factories have shown that automatic text classification has been a critical application, especially since the inception of digital documents (Hassan et al., 2022). The success of such techniques in the existing literature leads us to think that the implementation of machine learning techniques in educational research might be a valid means to identify teaching behaviours in a significantly more accurate and efficient way than the current approaches (self-report and record-and-code measures), through the training of supervised learning algorithms (classification) such as neural networks that, fed with verbal messages through speech-to-text techniques, allow behavioural patterns to be determined.

Improving efficiency in the identification and analysis of teaching behaviours through AI techniques might be the first step to the development of tools oriented to provide teachers with personalised feedback. When people are fully aware of their behaviours, they are more likely to implement behavioural changes. This is the case for PA tracking devices, for instance, which have been found to be effective to increase PA (Ferguson et al., 2022). In the educational domain, it has been suggested that when teachers are not fully aware of their own classroom practices, the effectiveness of educational interventions might be hindered (Sailer et al., 2023). Thus, moving towards automation would undoubtedly facilitate further development of personalised feedback solutions, which could implement intelligent recommender systems (IRS; de Schipper et al., 2021). This technology would provide a personalised set of recommendations for each teacher based on previous information (such as their teaching behaviours).

Findings from the present study are expected to have an important social impact on the different populations involved in the study. On the one hand, it is expected that PE teachers will benefit from the strategy proposal that will be made based on the findings obtained in the study. This product aims to enhance teachers' competences antd knowledge in dealing with students in their classes in a way such that their behaviours have a positive impact on students' adaptive motivational and behavioural patterns. Thus, the impact of the present research on PA-related motivational and behavioural patterns will take place indirectly, through the improvement of teaching behaviours. On the other hand, findings from the present project could be transferred to other populations such as teachers from other disciplines, or sport instructors and coaches, which would increase the social impact.

From a broader perspective, the study to be developed through this protocol seeks to generate interdisciplinary knowledge that can foster sustainable development through its contribution in addressing two sustainable development goals (SDG; Assembly, 2015). In this respect, it can contribute to SDG 3 (Good health and wellbeing) by providing teachers with useful tools and strategies to improve the interactions they have during PE lessons, which will consequently improve their personal and work development. It can also add to SDG 4 (Quality education), since it has been evidenced that teaching behaviours may influence different students' motivational and academic outcomes: thus, improving teacher-student interactions will enhance the teaching-learning process and therefore enrich the quality of education.

It is also important to acknowledge several technical and methodological challenges that could limit the immediate practical scope of this proposal. On the technical side, the quality of transcriptions in real PE settings represents a significant concern, since environments are often noisy, teacher commands are typically brief, and non-verbal expressivity plays a crucial role, which may compromise speech recognition accuracy. Moreover, traditional text representation techniques such as TF-IDF or Bag-of-Words, while useful in other contexts, might not be well suited to short and context-dependent utterances, as they risk losing essential semantic and pragmatic information. Similarly, classification models trained on generic embeddings may struggle to capture the communicative nuances that characterize teaching discourse in PE. To overcome these limitations, future work should explore the potential of contextualized language models such as BERT or Spanish-specific variants, which have shown superior performance in handling short conversational texts; fine-tuning such models on education-specific data could enhance the accurate classification of teacher behaviours. On the methodological side, certain behaviours may not be fully captured through voice recordings alone. For instance, some teacher messages could only be interpreted as controlling or not depending on accompanying facial gestures. To mitigate this issue during dataset creation, sessions will also be video recorded to provide complementary contextual information. In addition, it is possible that the developed algorithm may not achieve high levels of accuracy when distinguishing among multiple subdimensions of behaviour. Two contingency plans have been developed to address this: (i) reducing the number of categories by grouping eight subdimensions into four broader dimensions following the circumplex approach, and (ii) enlarging the dataset by incorporating additional teachers or more sessions per teacher. Recognizing these challenges and outlining concrete strategies to address them does not weaken the current proposal; on the contrary, it strengthens its methodological rigor and provides a clearer roadmap for future research.

Looking forward, AI could play a significant role in improving teaching practices in PE. By integrating AI technologies, real-time feedback could be provided to teachers during lessons. For example, AI could analyze vocal patterns, tone, and language use to assess how these factors influence student engagement and motivation. This real-time analysis could help educators make immediate adjustments to their approach, improving the overall quality of teaching. Furthermore, AI could enable personalized teaching strategies, analyzing students' responses to different teaching behaviors and providing customized recommendations based on individual needs.

AI could also be valuable in teacher training programs, helping educators refine their classroom management and instructional techniques in a virtual or simulated environment before they apply them in real classrooms. Such applications could revolutionize how teachers are trained and supported in their professional development. However, the implementation of AI in educational settings is not without challenges. One of the key concerns is the potential for AI to overlook important contextual cues, such as body language or facial expressions, which are crucial in PE settings. This underlines the importance of combining AI with traditional methods like observational analysis to ensure a complete understanding of teaching behaviors. Additionally, ethical considerations surrounding the use of AI must be carefully addressed. The collection and analysis of student data, even when anonymized, pose privacy and consent concerns, particularly when dealing with minors. Ensuring data security, transparency in consent procedures, and eliminating biases in AI systems will be critical for maintaining trust and protecting the rights of students and teachers.

While AI holds considerable potential for improving teaching and learning experiences, it must be introduced cautiously and responsibly. Teachers, students, and parents need to trust that AI systems will be used ethically and with respect for privacy. AI should complement human judgment, not replace it, ensuring that educators retain full control over their teaching practices. As AI technologies continue to evolve, it will be crucial to monitor their impact and refine the systems to align with the ethical standards and needs of educational settings.

Conclusions

The protocol study described in this manuscript is expected to be an important contribution to the advancement of knowledge about the PE context and its influence on health-related outcomes. The implementation of AI in the evaluation of teaching behaviours could have a massive impact on the current scientific approaches used to evaluate teaching performance, not only in PE, but also in other disciplines, as well as coaching behaviours in sport.

Financing

This study has been funded by the Government of Aragon for the development of I+D+i projects in priority and multidisciplinary lines for the period 2024-2026 (project code: PROY S01_24) and by Aristos Campus Mundus Research Projects Grant-Program 2024 (project code: ACM2024_17). Evelia Franco was supported by a Ramón y Cajal postdoctoral fellowship (RYC2022-036278-I) funded by the Spanish Ministry of Science an Innovation (MCIN), the State Research Agency (AEI), and the European Social Fund Plus (FSE+).

References

Aelterman, N., Vansteenkiste, M., Haerens, L., Soenens, B., Fontaine, J. R. J., & Reeve, J. (2019). Toward an integrative and fine-grained insight in motivating and demotivating teaching styles: The merits of a circumplex approach. Journal of Educational Psychology, 111(3), 497-521. https://doi.org/10.1037/edu0000293

Ahmadi, A., Noetel, M., Parker, P. D., Ryan, R. M., Ntoumanis, N., Reeve, J., Beauchamp, M. R., Dicke, T., Yeung, A., Ahmadi, M., Bartholomew, K. J., Chiu, T. K. F., Curran, T., Erturan, G., Flunger, B., Frederick, C. M., Froiland, J. M., González-Cutre, D., Haerens, L., Jeno, L. M., Koka, A., Krijgsman, C., Langdon, J. L., White, R. L., Litalien, D., Lubans, D. R., Mahoney, J. W., Nalipay, M. J., Patall, E., Perlman, D., Quested, E., Schneider, S., Standage, M., Stroet, K., Tessier, D., Thogersen-Ntoumani, C., Tilga, H., Vasconcellos, D., & Londsdale, C. (2023). A Classification System for Teachers' Motivational Behaviours Recommended in Self-Determination Theory Interventions. Journal of Educational Psychology, 115(8), 1158-1176. https://doi.org/10.1037/edu0000783

American Psychological Association. (2002). Ethical principles of psychologists and code of conduct (Amended August 3, 2016). American Psychological Association.

Barber, M., & Mourshed, M. (2007). Cómo hicieron los sistemas educativos con mejor desempeño del mundo para alcanzar sus objetivos. McKinsey & Company.

Burgueño, R., Abós, A., Sevil-Serrano, J., Haerens, L., De Cocker, K., & García-González, L. (2023). A Circumplex Approach to (de)motivating Styles in Physical Education: Situations-In-School–Physical Education Questionnaire in Spanish Students, Pre-Service, and In-Service Teachers. Measurement in Physical Education and Exercise Science, 28(1), 86-108. https://doi.org/10.1080/1091367X.2023.2248098

Cádiz Chacón, P., Barrio Mateu, L. A., León Valladares, D., Hernández Sánchez, Álvaro, Milla Palma, M., & Sotomayor Fernández, M. (2021). Motivación contextual desde la autodeterminación en las clases de Educación Física. Retos, 41, 88–94. https://doi.org/10.47197/retos.v0i41.80998

Chen, B., Vansteenkiste, M., Beyers, W., Boone, L., Deci, E. L., Van der Kaap-Deeder, J., Duriez, B., Lens, W., Matos, L., Mouratidis, A., Ryan, R. M., Sheldon, K. M., Soenens, B., Van Petegem, S., & Verstuyf, J. (2015). Basic psychological need satisfaction, need frustration, and need strength across four cultures. Motivation and emotion, 39(2), 216-236. https://doi.org/10.1007/s11031-014-9450-1

Cheon, S. H., Reeve, J., & Marsh, H. W. (2023). Autonomy-Supportive Teaching Enhances Prosocial and Reduces Antisocial Behavior via Classroom Climate and Psychological Needs: A Multilevel Randomized Control Intervention. Journal of Sport & Exercise Psychology, 45(1), 26-40. https://doi.org/10.1123/jsep.2021-0337

Claudino, J. G., Capanema, D. d. O., de Souza, T. V., Serrão, J. C., Machado Pereira, A. C., & Nassis, G. P. J. S. m.-o. (2019). Current approaches to the use of artificial intelligence for injury risk assessment and performance prediction in team sports: a systematic review. Sports Medicine - Open, 5(1), 1-12. https://doi.org/10.1186/s40798-019-0202-3

Comisión Europea. (2017). Modernisation of Higher Education in Europe: Academic Staff - 2017. Eurydice Report. .

Cudney, E. A., Anderson, S., Beane, R., Furterer, S., Mohandas, L., & Laux, C. (2023). Using the voice of the student to identify perceptions of teaching effectiveness attributes: a pilot study. Quality Assurance in Education, 31(3), 485-503. https://doi.org/10.1108/QAE-10-2022-0187

de Schipper, E., Feskens, R., & Keuning, J. (2021). Personalized and Automated Feedback in Summative Assessment Using Recommender Systems [Original Research]. 6. https://doi.org/10.3389/feduc.2021.652070

Dudley, D., Mackenzie, E., Van Bergen, P., Cairney, J., & Barnett, L. (2022). What Drives Quality Physical Education? A Systematic Review and Meta-Analysis of Learning and Development Effects From Physical Education-Based Interventions [Systematic Review]. Frontiers in Psychology, 13. https://doi.org/10.3389/fpsyg.2022.799330

Ferguson, T., Olds, T., Curtis, R., Blake, H., Crozier, A. J., Dankiw, K., Dumuid, D., Kasai, D., O'Connor, E., Virgara, R., & Maher, C. (2022). Effectiveness of wearable activity trackers to increase physical activity and improve health: a systematic review of systematic reviews and meta-analyses. The Lancet Digital Health, 4(8), e615-e626. https://doi.org/10.1016/S2589-7500(22)00111-X

Finegan, E. (2011). Language. Its Structure and Use. Wadsworth.

Franco, E., Coterón, J., Gómez, V., & Spray, C. M. (2021). A person-centred approach to understanding dark-side antecedents and students' outcomes associated with physical education teachers' motivation. Psychology of Sport and Exercise, 57, 102021. https://doi.org/10.1016/j.psychsport.2021.102021

Gao, Z., Chen, S., Sun, H. C., Wen, X., & Xiang, P. (2018). Physical Activity in Children's Health and Condition. BioMed Research International, 2018. https://doi.org/10.1155/2018/8542403

Guthold, R., Stevens, G. A., Riley, L. M., & Bull, F. C. (2020). Global trends in insufficient physical activity among adolescents: a pooled analysis of 298 population-based surveys with 1.6 million participants. The Lancet Child & Adolescent Health, 4(1), 23-35. https://doi.org/10.1016/S2352-4642(19)30323-2

Hassan, S. U., Ahamed, J., & Ahmad, K. (2022). Analytics of machine learning-based algorithms for text classification. Sustainable Operations and Computers, 3, 238-248. https://doi.org/https://doi.org/10.1016/j.susoc.2022.03.001

Hein, V., Müür, M., & Koka, A. (2004). Intention to be physically active after school graduation and its relationship to three types of intrinsic motivation. European Physical Education Review, 10(1), 5-19. https://doi.org/10.1177/1356336X04040618

Hill, Y., Lomas, L., & MacGregor, J. (2003). Students' perceptions of quality in higher education. Quality Assurance in Education, 11(1), 15-20. https://doi.org/10.1108/09684880310462047

Jatnika, D., Bijaksana, M. A., & Suryani, A. A. (2019). Word2Vec Model Analysis for Semantic Similarities in English Words. Procedia Computer Science, 157, 160-167. https://doi.org/https://doi.org/10.1016/j.procs.2019.08.153

Jiang, H., Lu, Y., & Xue, J. (2016). Automatic soccer video event detection based on a deep neural network combined cnn and rnn. 2016 IEEE 28th International Conference on Tools with Artificial Intelligence (ICTAI).

Lamoneda Prieto, J., Matos Duarte, M., Palacio, E. S., & Fraile, J. (2024). Impacto de un programa de intervención basado en la teoría de la autodeterminación sobre las necesidades psicológicas básicas, la intención de ser físicamente activo y la satisfacción con la vida de estudiantes de secundaria: Estudio longitudinal. Retos, 56, 228–237. https://doi.org/10.47197/retos.v56.103825

Larson, R. B. (2019). Controlling social desirability bias. International Journal of Market Research, 61(5), 534-547. https://doi.org/10.1177/1470785318805305

Leão Pereira, A. F., & Lorente-Catalán, E. (2024). Educación Física de Calidad: Diseño y validación de una herramienta orientada a la reflexión e innovación en los procesos educativos. Retos, 51, 32–46. https://doi.org/10.47197/retos.v51.99745

Mamani-Ramos, A. A., Damian-Nuñez, E. F., Carpio-Vargas, E. E., Mujica-Bermúdez, I., Pérez-Reátegui, C. M., Botton-Estrada, L. M., Quisocala-Ramos, J. A., Quispe-Cruz, H., Cutimbo-Quispe, C. V., Rodriguez- Mamani, J. R., Palomino-Crisóstomo, R. P., Cutipa-Salluca, W. R., Tuero-Chirinos, K. F., Villanueva-Alvaro, N. S., & Lava- Gálvez, J. J. (2025). Principales variables que predicen la competencia motora: análisis con árboles de clasificación. Retos, 68, 318–330. https://doi.org/10.47197/retos.v68.113 239

Moreno, J. A., Moreno, R., & Cervelló, E. (2007). El autoconcepto físico como predictor de la intención de ser físicamente activo. Psicología y Salud, 17(2), 261-267. https://doi.org/10.25009/pys.v17i2.710

Mourshed, M., Chijioke, C., & Barber, M. (2012). Cómo continúan mejorando los sistemas educativos de mayor progreso en el mundo. McKinsey & Company.

Pennington, J., Socher, R., & Manning, C. D. (2014). Glove: Global vector for word representation.

Reeve, J. (2013). How students create motivationally supportive learning environments for themselves: The concept of agentic engagement. Journal of Educational Psychology, 105(3), 579-595. https://doi.org/10.1037/a0032690

Reyes-Rodríguez, A. D., Retamal, F., Ho, W., & López de D'Amico, R. (2025). Percepción de la calidad de la Educación Física en República Dominicana. Retos, 64, 416–430. https://doi.org/10.47197/retos.v64.110570

Ryan, R. M., & Deci, E. L. (2000). Intrinsic and Extrinsic Motivations: Classic Definitions and New Directions. Contemporary Educational Psychology, 25(1), 54-67. https://doi.org/https://doi.org/10.1006/ceps.1999.1020

Sailer, M., Bauer, E., Hofmann, R., Kiesewetter, J., Glas, J., Gurevych, I., & Fischer, F. (2023). Adaptive feedback from artificial neural networks facilitates pre-service teachers' diagnostic reasoning in simulation-based learning. Learning and Instruction, 83, 101620. https://doi.org/10.1016/j.learninstruc.2022.101620

Santos, I., Nedjah, N., & Mourelle, L. M. (2017). Sentiment analysis using convolutional neural network with fastText embeddings. 2017 IEEE Latin American Conference on Computational Intelligence,

Shen, B., McCaughtry, N., Martin, J., Fahlman, M., & Garn, A. (2012). Urban high-school girls' sense of relatedness and their engagement in Physical Education. Journal of Teaching in Physical Education, 31, 231-245. https://doi.org/10.1123/jtpe.31.3.231

Simón-Chico, L., González-Peño, A., Hernández-Cuadrado, E., & Franco, E. (2023). The Impact of a Challenge-Based Learning Experience in Physical Education on Studens' Motivation and Engagement. European Journal of Investigation in Health, Psychology and Education, 13(4), 684-700.

Tilga, H., Kalajas-Tilga, H., Hein, V., Raudsepp, L., & Koka, A. (2021). Perceived autonomy support from peers, parents, and physical education teachers as predictors of physical activity and health-related quality of life among adolescents—a one-year longitudinal study. Education Sciences, 11(9). https://doi.org/10.3390/educsci11090457

Tomczak, M., & Tomczak, E. (2014). The need to report effect size estimates revisited. An overview of some recommended measures of effect size. Trends in Sport Sciences, 21(1).

UNESCO. (2015). Educación física de calidad: guía para los responsables políticos.

Van Doren, N., De Cocker, K., Flamant, N., Compernolle, S., Vanderlinde, R., & Haerens, L. (2023). Observing physical education teachers' need-supportive and need-thwarting styles using a circumplex approach: how does it relate to student outcomes? Physical Education and Sport Pedagogy, 1-25. https://doi.org/10.1080/17408989.2023.2230256

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention is all you need. 31st Conference on Neural Information Processing Systems.,

Zamarripa, J., Rodríguez-Medellín, R., Pérez-Garcia, J. A., Otero-Saborido, F., & Delgado, M. (2020). Mexican Basic Psychological Need Satisfaction and Frustration Scale in Physical Education. Frontiers in Psychology, 11(253). https://doi.org/10.3389/fpsyg.2020.00253

Zhang, Y., Jin, R., & Zhou, Z.-H. (2010). Understanding bag-of-words model: a statistical framework. International Journal of Machine Learning and Cybernetics, 1(1), 43-52. https://doi.org/10.1007/s13042-010-0001-0

Authors' and translators' details:

Evelia Franco	efranco@uloyola.es	Author and translator
Daniel Gutiérrez-Reina	dgutierrezreina@us.es	Author
Alba González-Peño	alba.gonzalez.peno@upm.es	Author
Javier Coterón	j.coteron@upm.es	Author