

# Academic achievement prediction in secondary education by decision tree analysis

## *Predicción del rendimiento académico en educación secundaria mediante el análisis de árboles de decisión*

Israel Villarrasa-Sapiña <sup>1</sup> 

Xavier García-Massó <sup>2\*</sup> 

Encarnación Liébana <sup>2</sup> 

Gonzalo Monfort Torres <sup>4</sup> 

<sup>1</sup> Universidad Internacional de La Rioja, Spain

<sup>2</sup> Universitat de València, Spain

<sup>3</sup> Universitat Catòlica de València, Spain

<sup>4</sup> Universitat Politècnica de València, Spain

\* Corresponding author. E-mail: [xavier.garcia@uv.es](mailto:xavier.garcia@uv.es)

### How to reference this article:

Villarrasa-Sapiña, I., García-Massó, X., Liébana, E., & Monfort Torres, G. (2024). Academic achievement prediction in secondary school by decision tree analysis. *Educación XX1*, 27(1), 253-279. <https://doi.org/10.5944/educxx1.33351>

**Date of received:** 08/03/2022

**Date of acceptance:** 17/04/2023

**Published online:** 02/01/2024

### ABSTRACT

The aim of the present study was to develop a predictive model of academic achievement (school success or failure) by applying a decision tree analysis. A cross-sectional study was carried out to design a system for the early detection of academic failure. 219 adolescents (aged 14 to 16) participated and information on their socioeconomic status, body mass index (BMI) percentile, physical activity, leisure time spent in front of screens, enjoyment,

hope, anger, anxiety, boredom, behavioral engagement, emotional engagement, cognitive engagement, self-perceived school performance and intention to go to university was collected as input variables in decision tree analysis. 6 failure and 3 success groups were found able to predict academic performance. Good accuracy was obtained in the training (80.11 %) and validation (81.40 %) datasets of the decision tree. It is possible to predict academic failure or success by assessing weight status, physical activity, anger and hope during school attendance, intention to go to university and self-perceived school performance.

**Keywords:** high schools, academic achievement, prediction, physical activity level, decision tree

## RESUMEN

El objetivo del presente estudio fue desarrollar un modelo de predicción del rendimiento académico (éxito o fracaso escolar) mediante la aplicación de un análisis de árbol de decisión. Se realizó un estudio transversal para diseñar un sistema de detección temprana del fracaso escolar. Participaron 219 adolescentes (de 14 a 16 años) y se recabó información de su estatus socioeconómico, percentil de índice de masa corporal (IMC), actividad física, tiempo de ocio frente a pantallas, niveles de disfrute, esperanza, ira, ansiedad, aburrimiento, compromiso conductual, compromiso emocional, compromiso cognitivo, rendimiento escolar autopercebido e intención de ir a la universidad, como variables de entrada en el análisis del árbol de decisión. Se encontraron 6 grupos de fracaso y 3 de éxito capaces de predecir el rendimiento académico. Se obtuvo una buena precisión en los conjuntos de datos de entrenamiento (80.11 %) y validación (81.40 %) del árbol de decisión. Es posible predecir el fracaso o el éxito académico mediante la evaluación del estado de peso, la actividad física, la ira y la esperanza durante la asistencia a la escuela, la intención de ir a la universidad y el rendimiento escolar autopercebido.

**Palabras clave:** educación secundaria, rendimiento académico, predicción, nivel de actividad física, árbol de decisión

## INTRODUCTION

School failure is a polysemic term often associated with not achieving an academic goal, which usually means not passing certain subjects or achieving a minimum degree. Without a doubt, being able to prevent such situations would reduce the students and families' frustration and would be a great advance for society, so that all students could have a good education (Alexander et al., 1997, 2001; Cairns et al., 1989). To academic failures, several authors have proposed different forms of early detection based on various factors, such as emotions, physical fitness, sedentary lifestyle or academic commitment, which are detailed below (Alzina & Escoda, 2012; D'Mello et al., 2008; Pekrun et al., 2002; Weiner, 1982). It should be noted that these forms of early detection are created to detect school failure early enough so that professionals in the education system can intervene and improve the student's situation.

A factor that has been shown to be highly relevant in predicting both academic performance and school failure is the socio-economic status of students' families (Parr & Bonitz, 2015; Trujillo-Torres et al., 2020). Concretely, school failure in Spain is not distributed equally in the socioeconomic stratification, since the student's social class can affect school failure and performance, since the percentage of school failure is higher in working class than in middle class children (Martínez-García, 2011).

Pekrun et al. (2002) defined student emotions as the students' personal experience when performing academic activities, identified them as a very important part of their personal motivation in achieving academic success and avoiding school failure. Along the same lines, D'Mello et al. (2008) highlighted the fact that knowing students' emotions was important to carry out a good teaching process, due to the links between cognition and emotion. These relationships have been described by explaining that students experience confusion when they face obstacles in their objectives, or detect contradictions, incongruities or anomalies in the teaching process (Festinger, 1962; Graesser & Olde, 2003). If confusion is not resolved it can lead to irritation, frustration, anger and sometimes even rage. It is therefore understandable that multiple studies have concluded that negative emotions such as anger, anxiety and boredom are negatively correlated with academic performance (Pekrun, 2006; Pekrun et al., 2011). However, too low values have also not been found to be positive for emotions such as anger (Lane et al., 2005; Pekrun et al., 2011). On the other hand, a learner can experience a range of positive emotions (such as enjoyment) when challenges are faced, knowledge is uncovered, and concepts are mastered. Students who are actively engaged in the learning process can have a flow-like experience, when they are so engrossed in the material that time and fatigue disappear (Csikszentmihalyi, 2014). In fact, a

positive emotion such as hope is considered to have sufficient potential to redirect underachieving students (Dixson, 2019).

Another factor that has been linked to academic performance is the practice of physical activity (PA) and some related concepts such as sedentary behavior or physical fitness. In a longitudinal study, Pellicer-Chenoll et al. (2015) concluded that the cluster of students with higher PA and fitness had a lower body mass index (BMI) and higher academic performance compared to classmates who performed less PA. In turn, the cluster of students with the lower PA showed lower levels of physical fitness, higher BMI and lower academic performance than the rest of the student's profiles. Several studies have found this kind of relationship between PA (Marques et al., 2017; Morales, Pellicer-Chenoll, et al., 2011; Rasberry et al., 2011; A. Singh et al., 2012; Sullivan et al., 2017) or physical fitness (Coe et al., 2013; Van Dusen et al., 2011; Wittberg et al., 2009) and academic performance. However, there are even studies that argue that there is no conclusive evidence on the beneficial effects of PA on students' overall cognitive and academic performance (Rasberry et al., 2011). These discrepancies between the different studies may be due to different conceptions of academic performance and the way of measuring and considering the practice of PA. In general, the relationship between the two variables is considered positive or non-existent (Singh et al., 2019).

The possibility of sedentary habits having a negative influence on academic performance (apart from the fact that PA could have a positive influence) has also been explored. Peiró-Velert et al. (2014), observed the influence of the time spent in the sedentary use of screens (e.g., video games, mobile phones, television...) on academic performance. The results showed that there was an inversely proportional relationship between academic performance and the use of screens.

Student engagement has also been explored as a possible factor with an influence on academic failure. Carini et al. (2006), corroborate that student engagement is positively linked to desirable learning outcomes such as critical thinking and grades. Dogan (2015) analyzed this factor divided into three engagement dimensions: cognitive, behavioral and emotional. His results showed that cognitive engagement predicted academic performance, but emotional and behavioral engagement were not predictors. Other studies mentioned that behavioral engagement is important to achieve positive academic results and prevent dropping out (Connell & Wellborn, 1991; Finn, 1989).

As explained above, it is now known that these factors in isolation have an impact on academic performance. However, a limited number of studies have been published with the aim of developing an early detection system of academic failure using most of the above factors.

Casillas et al. (2012), examined the combined effects of predictor variables on estimating academic failure. His findings highlight the importance of using

several predictor factors (i.e., psychosocial, and behavioral) to achieve an accurate estimation of students at risk of drop out.

Davis et al. (2014), conducted a study that assessed the extent to which several social-emotional skills learning (academic self-efficacy motivation, social connections, importance of school and school management, management of psychological and emotional distress, and academic stress) could be used as predictors of academic outcomes. Their results indicated that the combination of social-emotional learning subscales effectively discriminated between students who made positive progress toward high school graduation and those identified as having dropped out.

Zhang et al. (2018), focused their efforts on finding out the type of predictive model that achieved the best accuracy. They made a comparison between classification models, such as naive Bayes, support vector machines, decision tree and multilayer perceptron. They obtained better results with the last two. In fact, other studies have used decision trees as the analysis method to predict school drop-outs and have found results with a relatively high accuracy in their estimations (Quadri & Kalyankar, 2010; Veitch, 2004). It should be noted that this analysis for this type of study is highly relevant and important compared to other predictive analyses due to its efficacy and multiple benefits. For example, decision tree can handle different kinds of input data (i.e., nominal, numeric, and text), it is easy to understand, and it can process erroneous data set values, among others (Rokach & Maimon, 2014).

As a summary of the factors most frequently used to predict academic failure or success, we refer to one of the most recent reviews, conducted by Alyahyan and Düşteğör (2020), which provides information on the variables most commonly used in this type of study, based on York's (2015) (York et al., 2015) definition of academic success. According to this review, student demographics and psychological factors have proved to be two of the most widely used for prediction, along with prior academic achievement and students' environment factors.

As can be seen, although the theoretical basis seems to have a solid foundation on the most influential factors, other relevant factors such as PA, BMI or sedentary habits have not been addressed. There is, however, ample literature that has demonstrated their influence on academic achievement. This could be because the studies carried out with these types of variables have focused on examining academic performance rather than academic failure or pass. For this reason, it is quite possible that these factors have not appeared as relevant because they have not been analyzed.

As has been explained above, some studies have used relatively novel analysis methods (such as decision tree) to predict academic failure using combinations of psychological and demographic variables. However, no studies have been published

on developing early detection systems of academic failure using combinations of not only psychological and demographic characteristics but also variables related to lifestyles, like the practice of PA and sedentary activities. Therefore, we consider our work to be novel, as it includes the most influential factors on the prediction of students' academic success studied, according to Alyahyan and Düştegör (2020), so far plus other elements and aspects that had not been considered in an interrelated manner until now, as BMI, PA or sedentary habits. Furthermore, the application of a multifactorial and non-linear analysis, such as the decision tree, avoids the limitations that linear analyses may entail (e.g., reduction of statistical power when many factors are added or multicollinearity) and, in addition, this analysis provides classification and prediction results that can be easily interpreted visually. Last, the importance of this type of study lies in the fact that if a combination of factors that can identify school failure is found, in turn, actions can be promoted to help avoid them and thus seek to achieve success.

The aim of this study was therefore to develop a predictive model of academic achievement (school success or failure) by means of decision tree analysis, using emotions on attending to school, school engagement, PA, leisure time spent in sedentary activities that require the use of screens, socio-demographic characteristics, and school adjustment variables.

## **METHODS**

### **Study design and participants**

A cross-sectional study was carried out to design a system for the early detection of academic failure in students of third and fourth grades of secondary education in Spain. For this, students completed a set of questionnaires at the beginning of one quarter to measure PA, hours spent on screen sedentary activities, socioeconomic status, emotions related to class attendance, school engagement and school adjustment. These variables together with the BMI were used as input variables to design a classification tree to predict the academic results (success or failure) at the end of the quarter (output variable).

The sample was composed of 219 adolescents (aged 14 to 16). The participants were recruited from the compulsory secondary education schools in Valencia (Spain). The inclusion criteria were: i) should be between 14 and 16 years old (both inclusive), ii) not be neurologically or intellectually unable to understand and complete the questionnaires, and iii) refusal of their progenitors to participate in the study. The participants' characteristics are reported in Table 1.

**Table 1**  
*Students' characteristics*

Variable	Total sample (n=219)	Male (n=100)	Female (n=119)
Age (years)	14.94 (0.79)	15.05 (0.79)	14.84 (0.77)
Height (cm)	167.95 (8.82)	173.36 (7.96)	163.42 (6.73)
Weight (kg)	61.52 (11.16)	66.52 (11.91)	57.26 (8.49)
BMI percentile	61.93 (24.94)	63.77 (26.62)	60.38 (23.45)

*Note.* Data expressed as mean (standard deviation). BMI = Body Mass Index.

The procedures applied in this study were approved previously by the Institutional Review Board of the University of Valencia (Code: 1503291) while also meeting the requirements set out in the Declaration of Helsinki (1975, subsequently revised in 2008). The parents of the participants supplied their written informed consent before participating in the experiment.

## Measures

All the measures were taken at the participants' high school in their habitual classroom. The researchers explained how to complete the set of questionnaires and resolved the students' doubts. The time required to fill out all the questionnaires was between 50-60 minutes.

It should be noted that the questionnaires were administered at the beginning of the quarter. This allowed establishing a predictive relationship between the input variables and academic failure. The academic qualifications for each student obtained at the end of the quarter were anonymized. Their academic performance was codified as "suspended" if the student did not pass one of the subjects (academic failure) or "approved" if the student passed all the subjects in the quarter (academic success).

## Family Affluence Scale (FAS) II

This is a questionnaire to determine the socioeconomic status (SES) of the families of adolescents in European and North American countries (Currie et al.,

2008). Four objective questions were asked to quantify vehicles, vacation trips, personal bedroom, and computers, to estimate family wealth. The answers were codified from 0 (minimum number of vehicles, trips...) to 3 (maximum number of vehicles, trips...) depending on the number of possible responses of each item. The final score of the scale was computed as the mean value of the items' scores, 0 being the lowest SES and 2.25 the highest.

### **Body Mass Index Percentile**

Weight and height were self-reported by the participants and BMI (kg/m<sup>2</sup>) was calculated. It should be noted that Sherry et al. (2007), reported good validation results of the self-reported weight and height in adolescents. Growth tables (Kuczmarski et al., 2000) were also used to calculate the BMI percentile (adjusted for age and sex).

### **Physical Activity Questionnaire for Adolescents (PAQ-A)**

PAQ-A was first validated by Kowalski et al. (1997), as a modified version of the PA questionnaire for older children. They found a good convergent validity of this questionnaire in measuring the general PA level of high school students. Later, Martínez-Gómez et al. (2009), validated the Spanish version of the PAQ-A obtaining moderate relationships with accelerometer data ( $\rho = 0.34 - 0.39$ ). This questionnaire is useful for measuring PA levels from very low to very intense in the last 7 days and is appropriate for teenagers between the ages of 13 and 18. It consists of eight questions that assess different aspects of the PA performed by the adolescent in different periods of the day. PAQ-A is a simple questionnaire, easy to complete and manage in the school environment. The overall result of the test is a score of 1 to 5 points (1 the lowest value and 5 the highest value) to determine the level of PA performed by each teenager.

### **Adolescent Sedentary Activity Questionnaire (ASAQ)**

ASAQ is used to measure time spent in a range of sedentary behaviors outside school hours during a normal week (Hardy et al., 2007). In brief, in this questionnaire, participants answer questions about fifteen sedentary habits, with details of how long (hours and minutes) they carry out each one every week. For this study, only the seven items referring to leisure sedentary activities that required the use of a screen technology were used. The final score for the leisure time spent in sedentary

activities in front of screens was computed as the total time reported in each of these scores in minutes.

### **Emotional Scales Questionnaire Related to Class Attendance (AEQ)**

The AEQ was first designed by Pekrun et al. (2011), to measure the emotions experienced by students in relation to class attendance. The complete questionnaire consists of 24 scales measuring several emotions that are organized in three sections to assess class-related, learning-related, and test-related emotions. The items are answered based on a 5-point Likert scale, where the lowest value (1) refers to total disagreement with the statement, while the highest value (5) corresponds to complete agreement with what the item expresses. For this study, we selected the 42 items of the scales for enjoyment (8 items), anger (7 items), anxiety (11 items), boredom (9 items) and hope (7 items) during class assistance (class-related emotions) in the Spanish version. This version of the questionnaire was validated by Rosas (2015), who found good parameters regarding reliability and structural and construct validity. The final score for each subscale is computed as the mean value of its items.

### **School Engagement Measure Questionnaire (SEM)**

The School Engagement Measure questionnaire consists of 19 items, with a Likert format with a range of 5 points (Fredricks & McColskey, 2012). The Spanish version of the questionnaire was validated by Díaz et al., (2016), who determined that 16 items were clustered into three engagement subscales: behavioral (4 items; e.g., “I pay attention in class”), emotional (5 items; e.g., “I am interested in the work at school”) and cognitive (7 items; “When I read a book, I ask myself questions to make sure I understand what it is about”) engagement. The cognitive engagement refers to the level of taking part in school life and developing complex reasoning skills (Doğan, 2014). The concept of emotional engagement includes the student’s interest in school is accepted as the student’s reactions in the classroom and the student’s level of interest, boredom, unhappiness, happiness, and anxiety (Skinner et al., 1990). Finally, behavioral engagement is linked with participation in academic, social, or extracurricular activities. The score of each dimension is calculated as the mean value of the items assigned to that subscale.

## Brief Multidimensional School Adjustment Scale

The Brief Multidimensional School Adjustment Scale assesses the degree to which the adolescent is integrated into the school environment (Rubia et al., 2010). It consists of 10 items with a Likert format with a range of 6 points, which are divided into 3 dimensions: i) problems of adaptation to the school environment (items 6, 7, 8, 9 and 10), ii) self-perception of school performance (items 1, 2 and 5), and iii) intention to go to university (items 3 and 4). Of all the dimensions, only the self-perception of school performance and intention to go to university were used and computed as the mean of the items that make them up. It measures a positive and integrated self-concept as a student, as well as expectations of continuing with higher education (Rubia et al., 2010).

## Statistical Analysis

Data analysis was performed using the Matlab R2018a program (Mathworks Inc., Natick, USA). First, a classification tree was applied to obtain a prediction model of academic performance (i.e., students suspend some subjects) using as input variables SES, BMI percentile, PA, leisure time spent in front of screens, enjoyment, hope, anger, anxiety, boredom, behavioral engagement, emotional engagement, cognitive engagement, self-perceived school performance and intention to go to university.

The classification tree was validated using a subsample of the dataset. This technique consists of dividing the total number of available cases (i.e., 219) into two data sets: training (i.e., 80% of cases;  $n = 176$ ) and validation (i.e., 20% of cases;  $n = 43$ ). No significant differences were found between the training and validation datasets in the variables used in this study. The training dataset was then used to obtain the decision tree, while the validation dataset was used to verify its validity. The decision or classification tree is a method that divides the sample into two subgroups using an independent explanatory variable. For this, a cut-off point of the explanatory variable is established, which divides the sample into two sub-nodes based on the value of the subjects in the variable, i.e. the cases that are above the threshold from a group and the cases that are below form the other group. This process is repeated for each subgroup, until all the cases are correctly classified.

In this study, the CART algorithm with the deviance was used as the as the split criterion, which is a binary algorithm that divides each group into two subgroups. In addition, to avoid possible over-training (avoiding low external validity) a condition was laid down during the training process that each node should have at least 10 cases to reduce the final number of nodes and divisions. The classification

tree thus divides adolescents according to discriminating variables to classify all participants according to whether they passed all subjects (i.e., approved) or not (i.e., suspended).

Once the model was obtained, it was applied to the validation data set to obtain classification performance variables. The accuracy of the classification and the suspended and approved prognostic values were computed. These variables were calculated as described in Eqs. (1, 2 and 3).

$$\text{Accuracy (\%)} = \frac{\text{True Approved} + \text{True Suspended}}{\text{Total sample}} \cdot 100 \quad \text{Eq. 1}$$

$$\text{Suspended prognostic value (\%)} = \frac{\text{True Suspended}}{\text{True Suspended} + \text{False Suspended}} \cdot 100 \quad \text{Eq. 2}$$

$$\text{Approved prognostic value (\%)} = \frac{\text{True Approved}}{\text{True Approved} + \text{False Approved}} \cdot 100 \quad \text{Eq. 3}$$

## RESULTS

In the following section, the results obtained are provided with respect to the descriptive data of the sample, as well as those of the decision tree. Descriptive data are shown in Table 2.

**Table 2**  
*Descriptive data of the input variables of the classification tree*

Variable	Approved students	Suspended students	All Students
Socio-Economic Status	2.62 [2.53–2.72]	2.43 [2.35–2.52]	2.52 [2.45–2.58]
BMI percentile	58.43 [53.31–63.55]	64.66 [60.30–69.03]	61.93 [58.61–65.25]
Physical activity	2.43 [2.28–2.57]	2.38 [2.25–2.50]	2.40 [2.30–2.49]

Variable	Approved students	Suspended students	All Students
Leisure time spent in front of screens	315.40 [266.04–364.76]	367.73 [325.59–409.87]	344.79 [312.79–376.79]
Enjoyment	2.65 [2.48–2.81]	2.31 [2.15–2.46]	2.46 [2.34–2.57]
Hope	3.36 [3.16–3.57]	2.84 [2.65–3.02]	3.07 [2.93–3.21]
Anger	2.25 [2.10–2.40]	2.52 [2.33–2.71]	2.40 [2.28–2.53]
Anxiety	1.90 [1.72–2.07]	2.25 [2.07–2.43]	2.09 [1.97–2.22]
Boredom	2.75 [2.53–2.97]	2.99 [2.78–3.20]	2.88 [2.73–3.04]
Behavioral Engagement	3.98 [3.82–4.14]	3.49 [3.34–3.64]	3.70 [3.59–3.82]
Emotional Engagement	3.24 [3.04–3.44]	2.90 [2.74–3.07]	3.05 [2.92–3.18]
Cognitive Engagement	2.43 [2.27–2.59]	2.31 [2.18–2.44]	2.36 [2.26–2.46]
Self-perceived scholar performance	4.13 [3.93–4.33]	3.21 [3.02–3.40]	3.61 [3.46–3.76]
Intention to go to university	5.1 [4.78–5.42]	3.73 [3.37–4.09]	4.33 [4.07–4.59]

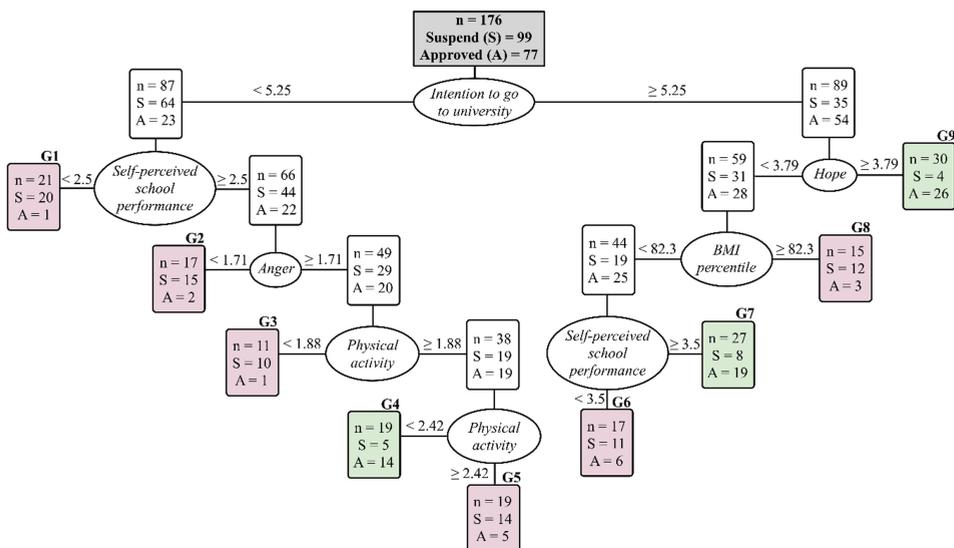
Note. Data expressed as mean [95 % confidence interval]. BMI = Body Mass Index.

The classification tree obtained with the training dataset is shown in Figure 1. The subjects assigned in this dataset are divided into subgroups using the input variables until the terminal nodes are reached. There were 9 terminal nodes, each representing a group of participants of the training data set. These terminal nodes are represented in Figure 1 as G1, G2, [...], G9 and were used to classify students based on academic failure (i.e., suspend) or success (i.e., pass).

As an example, the first node (i.e., G1) is explained below to clarify Figure 1. G1 consisted of 21 students, of whom 20 had suspended and 1 had passed the

evaluation. This is therefore considered an academic failure node with a participant grouping accuracy (from the training data set) of 95 % (i.e., 20 divided by 21). As the 21 students in this node presented an intention to go to the university lower than 5.25 points and a self-perceived school performance lower than 2.5 points, future students with these characteristics will have a high risk of school failure that should be reduced.

**Figure 1**  
Classification tree for adolescents in Approve (A) and Suspend (S) stages



Note. BMI = Body Mass Index.

Table 3 shows the characteristics of each terminal node to clarify the results. In the first column, each terminal node classifies students by academic success or failure, based on the proportion of people who pass or suspend from the training data set that were classified in each of them. The following columns show the variables in each terminal node to describe the characteristics of each one.

**Table 3***Characteristics of the terminal nodes of the decision tree*

	<b>BMIp</b>	<b>PA</b>	<b>Hope</b>	<b>Anger</b>	<b>SP</b>	<b>IGU</b>
G1 (Suspend)	–	–	–	–	<2.5	<5.25
G2 (Suspend)	–	–	–	<1.71	≥2.5	<5.25
G3 (Suspend)	–	<1.88	–	≥1.71	≥2.5	<5.25
G4 (Approve)	–	1.88–2.42	–	≥1.71	≥2.5	<5.25
G5 (Suspend)	–	≥2.42	–	≥1.71	≥2.5	<5.25
G6 (Suspend)	<82.3	–	<3.79	–	<3.5	≥5.25
G7 (Approve)	<82.3	–	<3.79	–	≥3.5	≥5.25
G8 (Suspend)	≥82.3	–	<3.79	–	–	≥5.25
G9 (Approve)	–	–	≥3.79	–	–	≥5.25

*Note.* body mass index percentile (BMIp), physical activity (PA), self-perceived scholar performance (SP), Intention to go to university (IGU).

Table 4 reported the performance of the decision tree to classify students from the training and validation data sets. From these results the decision tree was seen to perform well both in the training (80.11% accuracy) and validation (81.40% accuracy) datasets.

**Table 4***Performance variables of the decision tree in both training and validation data sets*

	<b>Decision tree</b>	
	<b>Training</b>	<b>Validation</b>
Accuracy (%)	80.11	81.40
APV (%)	77.63	79
SPV (%)	82	83.33

*Note.* Abbreviations: APV, Approved Prognostic Value; SPV, Suspended Prognostic Value.

## DISCUSSION AND CONCLUSIONS

The purpose of this manuscript was to develop an early detection system of academic failure by means of decision tree analysis using as potential input variables the emotions of students during class attendance, school engagement, intention to go to university, self-perceived school performance, socio-economic status, BMI percentile, PA and leisure time spent in sedentary activities that require the use of screens (e.g., watching TV). As supported by the results, the decision tree demonstrates a good classification accuracy as it was possible to implement this system in schools as an easy way of detecting adolescents at risk of suspending a subject. This system only requires the students to provide weight, height, PAQ-A (8 items), the items related with anger (7 items) and hope (7 items) during school attendance (i.e., from AEQ) as well as the items related with intention to go to university (2 items) and self-perceived school performance (3 items) from the brief multidimensional school adjustment scale. Therefore, although all the input variables used in this study are important factors in academic performance, not all of them are necessary to adequately predict school failure.

School failure is determined by the difficulties students encounter in reaching educational goals (Eisenberg et al., 2006; Enguita et al., 2010). As this is influenced by multiple factors, it is decisive to know which factors interfere to a greater or lesser extent in academic failure (Yu et al., 2018) to be able to develop early detection systems. To the authors knowledge, several studies have carried out an analysis focused on predicting academic failure (Casillas et al., 2012; Respondek et al., 2017; Yu et al., 2018). Casillas et al. (2012) use variables such as academic achievement, psychosocial characteristics and behavioral indicators, and Respondek et al. (2017) use perceived academic control and emotions (e.g., enjoyment, boredom and anxiety). Nevertheless, PA as well as the use of screen media technology during sedentary leisure activities have not been taken into account, despite the relationship found between these variables and academic performance (Morales, Pellicer-Chenoll, et al., 2011; Peiró-Velert et al., 2014).

By including all these variables in an analysis, a decision tree with high accuracy (i.e., 80.11 % in training data and 81.4 % in validation data) was obtained in our study. These results support previous studies, such as that of Zhang et al. (2018), who observed that the best methods to predict academic performance (five categories) were decision tree and multilayer perceptron (accuracy 57.41 and 62.04 % for the validation data set, respectively) as against Naïve Bayes or support vector machines (35.65 and 48.61 %, respectively). Vairachilai (2020), concluded that Naïve Bayes and decision tree had higher accuracy (77 and 71 %, respectively) than support vector machine (38 %). Finally, in the study by Ashraf et al. (2018), the decision tree was found to be the most accurate predictive analysis (97.3% accuracy). It should

be noted that Ashraf et al., did not apply a cross-validation procedure and their results (i.e., very high accuracy) could be indicative of an overtrained model. All in all, it can be said that the decision tree is a suitable tool for predicting academic performance and the performance parameters of our study are similar or higher than those presented in other studies with a similar classification purpose.

It should be noted that the decision tree obtained in this study determines the characteristics of the student groups that approved or suspended all the subjects. The results determined nine groups (terminal nodes of the decision tree), of which six represent groups of students that failed in some subject while the other three represent students who passed all the subjects.

Thus, there are six combinations of characteristics that lead to academic failure in adolescents. For example, students allocated in G1 (Figure 1) did not present a high intention to go to university and showed low self-perceived school performance, resulting in many probabilities (i.e., around 95 %) of suspending some of the subjects of the course (i.e., G1 in figure 1). But what is really interesting is that, on the left branch of the tree, the students with the intention of going to university lower than 5.25 need to have a relatively high score in self-perceived school performance ( $\geq 2.5$ ), to experience at least a minimum level of anger ( $\geq 1.88$ ) during some events attending school and to perform a moderate level of PA (from 1.88 to 2.42 points) to reach academic success (i.e., pass all subjects). Students with the intention of going to university lower than 5.25 who present any other combination of the above variables will have a high probability of failing a subject (i.e., G1, G2, G3 and G5 of figure 1). This fact is highly relevant for suggesting strategies to help these students, who have a moderate desire to go to university, to pass high school assessments. For example, if a student allocated in G3 is detected, it could be interesting to promote active lifestyles with the objective of moderately increasing the PA performed by the student in the hope of changing from G3 (failure) to G4 (success).

As can be seen in this decision tree, the intention to go to university is a variable of great importance, as it is necessary to know the rest of the factors that influence students. In contrast, in a previous study it had not been considered as an effective predictor variable (Fernandez-Lasarte et al., 2019). This may be due to the fact that this study has not analyzed this factor using non-linear analyses, which modulate the weight of each predictor variable by interrelating it with the other variables.

These results are in line with previous studies. On the one hand, it may seem contradictory that a very low value of anger is an important characteristic in determining school failure in some students (i.e., G2) since multiple studies mention that a high degree of anger is negative for academic performance (Pekrun, 2006; Pekrun et al., 2011). Nevertheless, some studies conducted, such as those by Pekrun et al. (2011) and Lane et al. (2005) concluded that a certain amount of anger can have positive effects on academic performance, while both very high and very low values

of anger are negative for students' academic achievement. This could be explained by the students' control of their emotional intelligence, since Parker et al. (2004) suggested that individuals with a high level of emotional intelligence are aware of the positive effects of anger on academic performance and are able to regulate their mood to reach appropriate states and achieve academic success. Therefore, if a student with a very low level of anger (i.e., allocated in G2) is detected, it would be interesting to analyze his or her ability in the emotional intelligence domain so that he or she can control emotions such as boredom (Pekrun et al., 2010), hopelessness (Titz, 2001) and anger, which, without proper control, can destabilize the emotional state and thus avoid school failure.

On the other hand, the results extracted from Groups 3, 4 and 5 corroborate that it is necessary for students to have a moderately active life in order to be academically successful. G3 (failure) reaffirms that lack of PA is positively correlated with school failure (Pellicer-Chenoll et al., 2015) and, inversely, G4 (success) corroborates that the moderate practice of PA is positively related to academic success (Morales, González, et al., 2011; Morales, Pellicer-Chenoll, et al., 2011; Pellicer-Chenoll et al., 2015). These results can be explained by neurophysiological reasons (Hillman et al., 2005; Tomporowski et al., 2008; van Praag, 2009), since physical exercise improves blood flow to the brain and thus cognitive functions are improved, or by psychosocial reasons (Sallis et al., 1999; Sigfúsdóttir et al., 2007), since PA is positively associated with mental health, self-esteem, emotional well-being and self-concept, which may have a positive influence on academic performance (Pellicer-Chenoll et al., 2015). Even so, this correlation between PA and academic performance is not entirely linear and Morales, Pellicer-Chenoll, et al. (2011) have already suggested that high levels of PA do not lead to improvements in academic performance, which would support the results found in G5 (failure). In other words, the relationship between PA and school performance could be non-linear, since the moderate levels are those most linked with school achievements (i.e., quadratic function).

Considering the right branch of the tree, it should be noted that there are two groups of characteristics that lead to academic success: i) intention to go to university equal or higher than 5.25 and experience hope during class attendance at least up to 3.79 points; and ii) intention to go to university equal or higher than 5.25, experience low-to-moderate levels of hope during class (i.e., <3.79), to have a BMI percentile lower than 82.3 and to have high self-perceived school performance (i.e.,  $\geq 3.5$ ). A good way of helping students motivated to go to university is to be sure that they experience hope during class attendance (i.e., help these students to be allocated in G9) to dramatically improve their academic success. These results are in line with other studies, which observed a positive correlation between hope and academic performance and proposed the application of interventions to increase academic hope due to its benefits (Feldman & Kubota, 2015).

However, if hope cannot be encouraged to high levels in some students, a healthy lifestyle should be promoted to control the BMI percentile and prevent students from becoming overweight (i.e., 85 ≤ 94 percentile) or obese (>95 percentile) (Kuczmariski et al., 2000) so that they do not fail academically (i.e., G8). These results are in line with previous studies in which having a higher percentage of BMI can cause disconnection or lack of commitment to academic work (Finn et al., 2018; Peterson et al., 2012). Those who are overweight or obese may have lower productivity due to health problems, which may be related to social, psychological and affective issues (Shaw et al., 2015). Since childhood overweight and obesity are related to PA, diet, and sedentary lifestyle, those students with healthy lifestyles could therefore have a lower BMI and probably achieve good academic performance (Pellicer-Chenoll et al., 2015). The results obtained in previous studies are therefore corroborated, confirming that obesity is important in determining academic performance and consequently that healthy lifestyles are also important factors. However, although healthy habits are important for academic success, the results of this study determine that the educational system should also try to make the students have a moderate-to-high self-perceived school performance in order to achieve academic success (i.e., G7) and avoid failure (i.e., G6), as shown in both branches of the tree. Therefore, if there is not high motivation to go to university and high hope feelings, then teachers should encourage this aspect more strongly by promoting moderate-to-high self-perceived performance in order to avoid academic failure. One way to improve self-perceived school performance could be to adjust the difficulty of classroom tasks to the level of the students, trying to find tasks that are difficult for them but at the same time accessible and passable. This would reduce stress and maintain psychological and emotional health (Solberg et al., 1998). Furthermore, this factor is even more relevant when observing the results of previous studies, where it has been found that a decrease in self-perceived school performance leads to a decrease in academic expectations (García-Escalera et al., 2020). This suggests that, for example, if a student is in G6, if his or her self-perceived school performance is not promoted, he or she could move to G1, increasing the probability of academic failure (see the ratio of failures and passes in these groups in Figure 1 to understand this example). In fact, Godoy et al., (2013) concluded that a negative academic self-perception can be a determining factor in academic performance, and they mention the need to work on students' self-perception at school because this could promote the improvement of academic performance (Moreira et al., 2016).

As has been demonstrated in the above paragraphs, knowing the characteristics of these groups would not only be interesting for understanding the nature of failing in the academic process, but also for implementing strategies to help students to be successful in school. The decision tree thus has two main practical applications:

to detect students at risk of academic failure and to provide an individualized orientation to design strategies to avoid failure.

A strong point of this study was its use of a decision tree approach for testing multifactorial combinations including variables related to active lifestyles. Also, it was carried out in only one city, so that its findings should be corroborated and extended to other places with a higher number of participants to increase the generalizability of the results. Another limitation is the limited number of input variables, since other interesting variables such as dietary habits or school social relationships could increase the performance of the decision tree in classifying students according to their academic results.

### Practical application

This predictive model of academic achievement presented in this study can have a practical transfer of knowledge to the education system, as it could be used as a tool to detect school failure. This tool would make it easier to guide the intervention of education professionals to improve their situation (i.e., teachers and specialist professionals). In order to implement it properly, school management should make these questionnaires available, as well as an explanation of their use. In addition, to further improve their usability, policy-makers could create a learning course or create a web application that explains, step by step, the procedure and provides feedback to address each detected situation of possible academic failure. In any case, a detailed explanation of how to apply this system with the tools currently available is presented below. These steps should only be followed for those students whose grades are low or who the teacher/professional feels may need it:

1. Answer the questions 3 and 4 of the *Brief Multidimensional School Adjustment Scale*. If the result is equal to or higher than 5.25 read point 2.a. and if it is lower than 5.25 read point 2.b.
- 2.a. Answer the 7 *hope* items of *Emotional Scales Questionnaire Related to Class Attendance (AEQ)*. If the result is lower than 3.79, the teacher would weigh and measure the student to calculate the BMI percentile. If the BMI percentile is equal or higher than 82.3, this student needs to be helped based on these characteristics. Instead, if the student has a lower percentile of 82.3, he/she must answer questions 1, 2 and 5 of the *Brief Multidimensional School Adjustment Scale*. If the result is lower than 3.5, this student also needs to be helped. In this way, with a maximum of 12 simple questions and the calculation of the BMI percentile, the academic achievement of the student and the reasons for his or her performance could be determined.

2.b. Answer the questions 1, 2 and 5 of the *Brief Multidimensional School Adjustment Scale*. If the result is lower than 2.5, this student needs to be helped based on these characteristics. However, if the result is equal or higher than 2.5, this student must answer the 7 *anger* items of *Emotional Scales Questionnaire Related to Class Attendance (AEQ)*. If the result is lower than 1.71, this student needs to be helped based on these characteristics. Last, if the result is equal or higher than 1.71, this student must answer the *Physical Activity Questionnaire for Adolescents (PAQ-A)* (8 items). If the result is lower than 1.88 or equal or higher than 2.42, this student needs to be helped based on these characteristics. In this way, with a maximum of 18 simple questions the academic achievement of the student and the reasons for his or her performance could be determined.

## REFERENCES

- Alexander, K. L., Entwisle, D. R., & Horsey, C. S. (1997). From first grade forward: early foundations of high school dropout. *Sociology of Education*, 70(2), 87–107. <https://doi.org/10.2307/2673158>
- Alexander, K. L., Entwisle, D. R., & Kabbani, N. S. (2001). The dropout process in life course perspective: early risk factors at home and school. *Teachers College Record*, 103(5), 760–822. <https://doi.org/10.1111/0161-4681.00134>
- Alyahyan, E., & Düşteğör, D. (2020). Predicting academic success in higher education: literature review and best practices. *International Journal of Educational Technology in Higher Education*, 17(1), 3. <https://doi.org/10.1186/s41239-020-0177-7>
- Alzina, R. B., & Escoda, N. P. (2012). Educación emocional: estrategias para su puesta en práctica. *Avances en Supervisión Educativa*, 16, 16. <https://avances.adide.org>
- Ashraf, A., Anwer, S., & Khan, M. G. (2018). A Comparative study of predicting student's performance by use of data mining techniques. *American Scientific Research Journal for Engineering, Technology, and Sciences*, 44(1), 1.
- Cairns, R. B., Cairns, B. D., & Neckerman, H. J. (1989). Early school dropout: configurations and determinants. *Child Development*, 60(6), 1437–1452. <https://doi.org/10.2307/1130933>
- Carini, R. M., Kuh, G. D., & Klein, S. P. (2006). Student engagement and student learning: testing the linkages. *Research in Higher Education*, 47(1), 1–32. <https://doi.org/10.1007/s11162-005-8150-9>
- Casillas, A., Robbins, S., Allen, J., Kuo, Y.-L., Hanson, M. A., & Schmeiser, C. (2012). Predicting early academic failure in high school from prior academic achievement, psychosocial characteristics, and behavior. *Journal of Educational Psychology*, 104(2), 407–420. <https://doi.org/10.1037/a0027180>

- Coe, D. P., Peterson, T., Blair, C., Schutten, M. C., & Peddie, H. (2013). Physical fitness, academic achievement, and socioeconomic status in school-aged youth. *The Journal of School Health, 83*(7), 500–507. <https://doi.org/10.1111/josh.12058>
- Connell, J. P., & Wellborn, J. G. (1991). Competence, autonomy, and relatedness: a motivational analysis of self-system processes. In M. R. Gunnar & L. A. Sroufe (Eds.) *Self processes and development* (pp. 43–77). Lawrence Erlbaum Associates, Inc.
- Csikszentmihalyi, M. (2014). Toward a psychology of optimal experience. In M. Csikszentmihalyi (Ed.), *Flow and the foundations of positive psychology: the collected works of mihaly csikszentmihalyi* (pp. 209–226). Springer Netherlands. [https://doi.org/10.1007/978-94-017-9088-8\\_14](https://doi.org/10.1007/978-94-017-9088-8_14)
- Currie, C., Molcho, M., Boyce, W., Holstein, B., Torsheim, T., & Richter, M. (2008). Researching health inequalities in adolescents: the development of the Health Behaviour in School-Aged Children (HBSC) family affluence scale. *Social Science & Medicine, 66*(6), 1429–1436. <https://doi.org/10.1016/j.socscimed.2007.11.024>
- Davis, A., Solberg, V. S., Baca, C. de, & Gore, T. H. (2014). Use of social emotional learning skills to predict future academic success and progress toward graduation. *Journal of Education for Students Placed at Risk, 19*(3–4), 169–182. <https://doi.org/10.1080/10824669.2014.972506>
- Díaz, E. R., Fernández, A. R., & Revuelta, L. R. (2016). Validation of the Spanish version of the School Engagement Measure (SEM). *The Spanish Journal of Psychology, 19*, 85–89.
- Dixson, D. D. (2019). Hope into action: how clusters of hope relate to success-oriented behavior in school. *Psychology in the Schools, 56*(9), 1493–1511. <https://doi.org/10.1002/pits.22299>
- D’Mello, S., Jackson, G., Craig, S., Morgan, B., Chip-Man, P., White, H., Person, N., Kort, B., Kaliouby, R., Picard, R., & Graesser, A. (2008). *AutoTutor detects and responds to learners affective and cognitive states*. Workshop on emotional and cognitive issues at the Int. Conf. Intelligent Tutoring Systems. Montreal, Canada.
- Doğan, U. (2014). Validity and reliability of Student Engagement Scale. *Bartın University Journal of Faculty of Education, 3*(2), 390-403. <https://doi.org/10.14686/BUFAD.201428190>
- Dogan, U. (2015). Student engagement, academic self-efficacy, and academic motivation as predictors of academic performance. *The Anthropologist, 20*(3), 553–561. <https://doi.org/10.1080/09720073.2015.11891759>
- Eisenberg, N., Damon, W., & Lerner, R. M. (2006). *Handbook of child psychology: social, emotional, and personality development (6th ed)*. John Wiley & Sons, Inc.
- Enguita, M. F., Martínez, L. M., & Gómez, J. R. (2010). *Fracaso y abandono escolar en España*. Fundación la Caixa.

- Feldman, D. B., & Kubota, M. (2015). Hope, self-efficacy, optimism, and academic achievement: distinguishing constructs and levels of specificity in predicting college grade-point average. *Learning and Individual Differences, 37*, 210–216. <https://doi.org/10.1016/j.lindif.2014.11.022>
- Fernandez-Lasarte, O., Goñi, E., Camino, I., & Zubeldia, M. (2019). Ajuste escolar y autoconcepto académico en la Educación Secundaria. *Revista de Investigación Educativa, 37*(1), 1. <https://doi.org/10.6018/rie.37.1.308651>
- Festinger, L. (1962). *A theory of Cognitive Dissonance*. Stanford University Press.
- Finn, J. D. (1989). Withdrawing from school. *Review of Educational Research, 59*(2), 117–142. <https://doi.org/10.3102/00346543059002117>
- Finn, K. E., Faith, M. S., & Seo, Y. S. (2018). School engagement in relation to body mass index and school achievement in a high-school age sample. *Journal of Obesity, 2018*, 3729318. <https://doi.org/10.1155/2018/3729318>
- Fredricks, J. A., & McColskey, W. (2012). The measurement of student engagement: a comparative analysis of various methods and student self-report instruments. In S. L. Christenson, A. L. Reschly, & C. Wylie (Eds.), *Handbook of research on student engagement* (pp. 763–782). Springer US. [https://doi.org/10.1007/978-1-4614-2018-7\\_37](https://doi.org/10.1007/978-1-4614-2018-7_37)
- García-Escalera, J., Valiente, R. M., Sandín, B., Ehrenreich-May, J., & Chorot, P. (2020). Los efectos de un programa de prevención de la ansiedad y la depresión para adolescentes en variables educativas y de bienestar. *Revista de Psicodidáctica, 25*(2), 143–149. <https://doi.org/10.1016/j.psicod.2020.05.001>
- Godoy, J. A., Abrahão, R. C., & Halpern, R. (2013). Autopercepção de dificuldades escolares em alunos do ensino fundamental e médio em município do Rio Grande do Sul. *Aletheia, 41*, 121–133.
- Graesser, A. C., & Olde, B. A. (2003). How does one know whether a person understands a device? The quality of the questions the person asks when the device breaks down. *Journal of Educational Psychology, 95*(3), 524–536. <https://doi.org/10.1037/0022-0663.95.3.524>
- Hardy, L. L., Booth, M. L., & Okely, A. D. (2007). The reliability of the Adolescent Sedentary Activity Questionnaire (ASAQ). *Preventive Medicine, 45*(1), 71–74. <https://doi.org/10.1016/j.ypmed.2007.03.014>
- Hillman, C. H., Castelli, D. M., & Buck, S. M. (2005). Aerobic fitness and neurocognitive function in healthy preadolescent children. *Medicine and Science in Sports and Exercise, 37*(11), 1967–1974. <https://doi.org/10.1249/01.mss.0000176680.79702.ce>
- Kowalski, K. C., Crocker, P. R. E., & Kowalski, N. P. (1997). Convergent validity of the Physical Activity Questionnaire for Adolescents. *Pediatric Exercise Science, 9*(4), 342–352. <https://doi.org/10.1123/pes.9.4.342>

- Kuczarski, R., Ogden, C., Grummer-Strawn, L., Flegal, K., Shumei, G., Wei, R., Mei, Z., Curtin, L., Roche, A., & Johnson, C. (2000). CDC growth charts: United States. *National Center for Health Statistics, 314*, 28.
- Lane, A. M., Whyte, G. P., Terry, P. C., & Nevill, A. M. (2005). Mood, self-set goals and examination performance: the moderating effect of depressed mood. *Personality and Individual Differences, 39*(1), 143–153. <https://doi.org/10.1016/j.paid.2004.12.015>
- Marques, A., Corrales, F. R. G., Martins, J., Catunda, R., & Sarmiento, H. (2017). Association between physical education, school-based physical activity, and academic performance: a systematic review. *Retos: Nuevas Tendencias En Educación Física, Deporte y Recreación, 31*, 316–320.
- Martínez García, J. S. (2011). Género y origen social: diferencias grandes en fracaso escolar administrativo y bajas en rendimiento educativo. *Revista de Sociología de la Educación-RASE, 4*(3), 270–282.
- Martínez-Gómez, D., Martínez-de-Haro, V., Pozo, T., Welk, G. J., Villagra, A., Calle, M. E., Marcos, A., & Veiga, O. L. (2009). Fiabilidad y validez del cuestionario de actividad física PAQ-A en adolescentes españoles. *Revista Española de Salud Pública, 83*(3), 427–439.
- Morales, J., González, L.-M., Guerra, M., Virgili, C., & Unnithan, V. (2011). Physical activity, perceptual-motor performance, and academic learning in 9-to-16-years-old school children. *International Journal of Sport Psychology, 42*(4), 401–415.
- Morales, J., Pellicer-Chenoll, M., García-Masso, X., Gomis, M., & González, L.-M. (2011). Relation between physical activity and academic performance in 3rd-year secondary education students. *Perceptual and Motor Skills, 113*(2), 539–546.
- Moreira, B. B. G., Martins-Reis, V. de O., & Santos, J. N. (2016). Autopercepção das dificuldades de aprendizagem de estudantes do ensino fundamental. *Audiology — Communication Research, 21*. <https://doi.org/10.1590/2317-6431-2015-1632>
- Parker, J. D. A., Summerfeldt, L. J., Hogan, M. J., & Majeski, S. A. (2004). Emotional intelligence and academic success: examining the transition from high school to university. *Personality and Individual Differences, 36*(1), 163–172. [https://doi.org/10.1016/S0191-8869\(03\)00076-X](https://doi.org/10.1016/S0191-8869(03)00076-X)
- Parr, A. K., & Bonitz, V. S. (2015). Role of family background, student behaviors, and school-related beliefs in predicting high school dropout. *The Journal of Educational Research, 108*(6), 504–514. <https://doi.org/10.1080/00220671.2014.917256>
- Peiró-Velert, C., Valencia-Peris, A., González, L. M., García-Massó, X., Serra-Añó, P., & Devís-Devís, J. (2014). Screen media usage, sleep time and academic

- performance in adolescents: clustering a self-organizing maps analysis. *Plos One*, 9(6), e99478. <https://doi.org/10.1371/journal.pone.0099478>
- Pekrun, R. (2006). The control-value theory of achievement emotions: assumptions, corollaries, and implications for educational research and practice. *Educational Psychology Review*, 18(4), 315–341. <https://doi.org/10.1007/s10648-006-9029-9>
- Pekrun, R., Goetz, T., Daniels, L. M., Stupnisky, R. H., & Perry, R. P. (2010). Boredom in achievement settings: exploring control–value antecedents and performance outcomes of a neglected emotion. *Journal of Educational Psychology*, 102(3), 531–549. <https://doi.org/10.1037/a0019243>
- Pekrun, R., Goetz, T., Frenzel, A. C., Barchfeld, P., & Perry, R. P. (2011). Measuring emotions in students' learning and performance: the Achievement Emotions Questionnaire (AEQ). *Contemporary Educational Psychology*, 36(1), 36–48. <https://doi.org/10.1016/j.cedpsych.2010.10.002>
- Pekrun, R., Goetz, T., Titz, W., & Perry, R. (2002). Academic emotions in students' self-regulated learning and achievement: a program of qualitative and quantitative research. *Educational Psychologist*, 37, 91–105. [https://doi.org/10.1207/S15326985EP3702\\_4](https://doi.org/10.1207/S15326985EP3702_4)
- Pellicer-Chenoll, M., Garcia-Massó, X., Morales, J., Serra-Añó, P., Solana-Tramunt, M., González, L.-M., & Toca-Herrera, J.-L. (2015). Physical activity, physical fitness and academic achievement in adolescents: a self-organizing maps approach. *Health Education Research*, 30(3), 436–448. <https://doi.org/10.1093/her/cyv016>
- Peterson, J. L., Puhl, R. M., & Luedicke, J. (2012). An experimental assessment of physical educators' expectations and attitudes: the importance of student weight and gender. *Journal of School Health*, 82(9), 432–440. <https://doi.org/10.1111/j.1746-1561.2012.00719.x>
- Quadri, M. M. N., & Kalyankar, D. N. V. (2010). Drop out feature of student data for academic performance using decision tree techniques. *Global Journal of Computer Science and Technology*, 2(10). [https://globaljournals.org/GJCST\\_Volume10/gjcst\\_vol10\\_issue2\\_ver1\\_paper7.pdf](https://globaljournals.org/GJCST_Volume10/gjcst_vol10_issue2_ver1_paper7.pdf)
- Rasberry, C. N., Lee, S. M., Robin, L., Laris, B. A., Russell, L. A., Coyle, K. K., & Nihiser, A. J. (2011). The association between school-based physical activity, including physical education, and academic performance: a systematic review of the literature. *Preventive Medicine*, 52, S10–S20. <https://doi.org/10.1016/j.ypmed.2011.01.027>
- Respondek, L., Seufert, T., Stupnisky, R., & Nett, U. E. (2017). Perceived academic control and academic emotions predict undergraduate university student success: examining effects on dropout intention and achievement. *Frontiers in Psychology*, 8. <https://doi.org/10.3389/fpsyg.2017.00243>

- Rokach, L., & Maimon, O. (2014). *Data mining with decision trees: theory and applications (2<sup>a</sup> ed.)*. World Scientific Publishing Co., Inc.
- Rosas, J. S. (2015). The Achievement Emotions Questionnaire-Argentine (AEQ-AR): internal and external validity, reliability, gender differences and norm-referenced interpretation of test scores. *Revista Evaluar*, 15(1), 1. <https://doi.org/10.35670/1667-4545.v15.n1.14908>
- Rubia, J. C. M. de la, Sosa, J. C. S., & González, M. E. V. (2010). Desarrollo de una escala multidimensional breve de ajuste escolar. *Revista Electrónica de Metodología Aplicada*, 15(1), 1. <https://doi.org/10.17811/rema.15.1.2010.1-11>
- Sallis, J. F., McKenzie, T. L., Kolody, B., Lewis, M., Marshall, S., & Rosengard, P. (1999). Effects of health-related physical education on academic achievement: Project SPARK. *Research Quarterly for Exercise and Sport*, 70(2), 127–134. <https://doi.org/10.1080/02701367.1999.10608030>
- Shaw, S. R., Gomes, P., Polotskaia, A., & Jankowska, A. M. (2015). The relationship between student health and academic performance: implications for school psychologists. *School Psychology International*, 36(2), 115–134. <https://doi.org/10.1177/0143034314565425>
- Sherry, B., Jefferds, M. E., & Grummer-Strawn, L. M. (2007). Accuracy of adolescent self-report of height and weight in assessing overweight status: a literature review. *Archives of Pediatrics & Adolescent Medicine*, 161(12), 1154–1161. <https://doi.org/10.1001/archpedi.161.12.1154>
- Sigfúsdóttir, I. D., Kristjánsson, A. L., & Allegrante, J. P. (2007). Health behaviour and academic achievement in Icelandic school children. *Health Education Research*, 22(1), 70–80. <https://doi.org/10.1093/her/cyl044>
- Singh, A. S., Saliasi, E., Berg, V. van den, Uijtdewilligen, L., Groot, R. H. M. de, Jolles, J., Andersen, L. B., Bailey, R., Chang, Y.-K., Diamond, A., Ericsson, I., Etnier, J. L., Fedewa, A. L., Hillman, C. H., McMorris, T., Pesce, C., Pühse, U., Tomporowski, P. D., & Chinapaw, M. J. M. (2019). Effects of physical activity interventions on cognitive and academic performance in children and adolescents: a novel combination of a systematic review and recommendations from an expert panel. *British Journal of Sports Medicine*, 53(10), 640–647. <https://doi.org/10.1136/bjsports-2017-098136>
- Singh, A., Uijtdewilligen, L., Twisk, J. W. R., van Mechelen, W., & Chinapaw, M. J. M. (2012). Physical activity and performance at school: a systematic review of the literature including a methodological quality assessment. *Archives of Pediatrics & Adolescent Medicine*, 166(1), 49–55. <https://doi.org/10.1001/archpediatrics.2011.716>
- Skinner, E. A., Wellborn, J. G., & Connell, J. P. (1990). What it takes to do well in school and whether I've got it: a process model of perceived control and children's

- engagement and achievement in school. *Journal of Educational Psychology*, 82(1), 22–32. <https://doi.org/10.1037/0022-0663.82.1.22>
- Solberg, V. S., Gusavac, N., Hamann, T., Felch, J., Johnson, J., Lamborn, S., & Torres, J. (1998). The Adaptive Success Identity Plan (ASIP): a career intervention for college students. *The Career Development Quarterly*, 47(1), 48–95. <https://doi.org/10.1002/j.2161-0045.1998.tb00728.x>
- Sullivan, R. A., Kuzel, A. H., Vaandering, M. E., & Chen, W. (2017). The association of physical activity and academic behavior: a systematic review. *Journal of School Health*, 87(5), 388–398. <https://doi.org/10.1111/josh.12502>
- Titz, W. (2001). *Emotionen von Studierenden in Lernsituationen: explorative Analysen und Entwicklung von Selbstberichtskalen — anhang*. <https://epub.uni-regensburg.de/9862/>
- Tomprowski, P. D., Davis, C. L., Miller, P. H., & Naglieri, J. A. (2008). Exercise and children's intelligence, cognition, and academic achievement. *Educational Psychology Review*, 20(2), 111–131. <https://doi.org/10.1007/s10648-007-9057-0>
- Trujillo-Torres, J.-M., Hossein-Mohand, H., Gómez-García, M., Hossein-Mohand, H., & Hinojo-Lucena, F.-J. (2020). Estimating the academic performance of secondary education mathematics students: a gain lift predictive model. *Mathematics*, 8(12), 12. <https://doi.org/10.3390/math8122101>
- Vairachilai S, V. (2020). Student's academic performance prediction using machine learning approach. *International Journal of Advanced Science and Technology*, 29, 9s.
- Van Dusen, D. P., Kelder, S. H., Kohl, H. W., Ranjit, N., & Perry, C. L. (2011). Associations of physical fitness and academic performance among schoolchildren. *The Journal of School Health*, 81(12), 733–740. <https://doi.org/10.1111/j.1746-1561.2011.00652.x>
- van Praag, H. (2009). Exercise and the brain: something to chew on. *Trends in Neurosciences*, 32(5), 283–290. <https://doi.org/10.1016/j.tins.2008.12.007>
- Veitch, W. R. (11-14 de abril de 2004). *Identifying characteristics of high school dropouts: data mining with a decision tree model* [Ponencia de Congreso]. Annual Meeting of the American Educational Research Association, San Diego. <https://eric.ed.gov/?id=ED490086>
- Weiner, B. (1982). An attribution theory of motivation and emotion. *Series in Clinical & Community Psychology: Achievement, Stress, & Anxiety*, 223–245.
- Wittberg, R. A., Northrup, K. L., & Cottrel, L. (2009). Children's physical fitness and academic performance. *American Journal of Health Education*, 40(1), 30–36.
- York, T. T., Gibson, C., & Rankin, S. (2015). Defining and measuring academic success. *Practical Assessment, Research & Evaluation*, 20(5). <https://eric.ed.gov/?id=EJ1059739>

- Yu, L. C., Lee, C. W., Pan, H. I., Chou, C. Y., Chao, P. Y., Chen, Z. H., Tseng, S. F., Chan, C. L., & Lai, K. R. (2018). Improving early prediction of academic failure using sentiment analysis on self-evaluated comments. *Journal of Computer Assisted Learning*, 34(4), 358–365. <https://doi.org/10.1111/jcal.12247>
- Zhang, X., Xue, R., Liu, B., Lu, W., & Zhang, Y. (28-30 de julio de 2018). *Grade prediction of student academic performance with multiple classification models*. [Ponencia de Congreso]. 14th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD), China. <https://doi.org/10.1109/FSKD.2018.8687286>

