KU LEUVEN

FACULTY OF PSYCHOLOGY AND
EDUCATIONAL SCIENCES

CENTER FOR PROFESSIONAL LEARNING & DEVELOPMENT,
CORPORATE TRAINING AND LIFELONG LEARNING

An artificial neural network analysis of academic perfomance in higher education

Carlos Felipe Rodríguez Hernández

Doctoral thesis offered to obtain the degree of
Doctor of Educational Sciences (PhD)

Supervisor: Prof. dr. Eva Kyndt
Co-Supervisor: Prof. dr. Eduardo Cascallar

2020

An artificial neural network analysis of academic performance in higher education

Dissertation submitted to obtain the degree of Doctor of Educational Sciences

Presented by: Carlos Felipe Rodríguez Hernández

Supervisor: Prof. Dr. Eva Kyndt Co-supervisor: Prof. Dr. Eduardo Cascallar

---

The understanding of the influential factors of academic performance in higher education contributes to improving its quality. Educational researchers have usually relied on the use of traditional statistical techniques to reach that understanding, possibly ignoring the complex relationship between academic performance and its predictors. The emerging field of educational data mining (EDM) has shown that artificial neural networks are a powerful tool to overcome this limitation, mainly when large amounts of data are analyzed. Nevertheless, due to a general lack of knowledge, artificial neural networks remain unknown and unexplored among educational researchers. Therefore, this doctoral dissertation aims to extend both conceptually and methodologically the use of artificial neural networks to analyze academic performance in higher education.

This doctoral dissertation consists of an introduction, followed by one systematic literature review, three empirical studies, and a discussion chapter. In Chapter 2, the relationship between SES and academic performance in higher education is untangled by critically reviewing the existing educational literature on this topic. Next, three empirical studies frame the use of artificial neural networks by classifying, predicting, and clustering academic performance in higher education. To do so, a unique dataset coming from the Colombian educational setting and provided by the Colombian Institute for Educational Evaluation is analyzed. This dataset contains information on academic performance in higher education, socio-economic status, prior academic achievement, high school characteristics, and working status of a cohort of 162,030 Colombian students.

In Chapter 3, multilayer perceptron (MLPs) are systematically designed to classify academic performance in higher education. Findings showed that it is possible to accurately classify academic performance as high or low, to assess the quality of the classification through several quality measures, and to establish the predictive contribution of the investigated predictors. In Chapter 4, radial basis function neural networks (RBFNNs) make it possible to predict academic performance in an accurate and interpretable way. Notably, it was found that prior academic achievement is strongly related to academic performance in higher education, regardless of the influence of SES on prior academic achievement. In Chapter 5, self-organizing maps (SOMs) are used to cluster students based on the investigated predictors. Interesting patterns among prior academic achievement, SES, and academic performance emerged from the analysis. Furthermore, three statistically different groups of students were identified, namely high, medium, and low. The final chapter of this doctoral dissertation summarizes the main findings of this research, describes its main contributions to the field of educational sciences, and sketches ideas for the future use of artificial neural networks in educational research.

Een artificiële neurale netwerkanalyse van academische prestaties in het hoger onderwijs

Dissertatie aangeboden tot het verkrijgen van de graad van Doctor in de Pedagogische Wetenschappen

Door: Carlos Felipe Rodríguez Hernández

Promotor: Prof. Dr. Eva Kyndt Co-promotor: Prof. Dr. Eduardo Cascallar

---

Door een sterker begrip te ontwikkelen van factoren die academische prestaties beïnvloeden, kan van de kwaliteit van het hoger onderwijs verbeterd worden. Hiervoor wordt doorgaans gebruik gemaakt traditionele statistische technieken, waarbij onderwijsonderzoekers mogelijk de complexe relatie tussen academische prestaties en de voorspellers ervan negeren. Het opkomende gebied van onderwijskundige *datamining* (EDM) heeft aangetoond dat artificiële neurale netwerken een krachtig hulpmiddel zijn om tegemoet te komen aan de beperkingen van traditionele technieken, vooral wanneer grote hoeveelheden gegevens worden geanalyseerd. Vanwege een algemeen gebrek aan kennis blijven deze artificiële neurale netwerken echter onbekend bij onderwijsonderzoekers. Dit proefschrift heeft als doel om het gebruik van artificiële neurale netwerken uit te breiden om academische prestaties in het hoger onderwijs te analyseren, zowel op conceptueel als op methodologisch vlak.

Dit proefschrift bestaat uit een inleiding, gevolgd door een systematisch literatuuronderzoek, drie empirische studies, waarna een algemene discussie volgt. Hoofdstuk 2 omvat een mixed method systematische review studie die inzicht biedt in de relatie tussen SES en academische prestaties in het hoger onderwijs. Vervolgens kaderen drie empirische studies het gebruik van artificiële neurale netwerken door academische prestaties in het hoger onderwijs te classificeren, voorspellen en clusteren. Om dit te doen, wordt een unieke dataset geanalyseerd die afkomstig is uit de Colombiaanse onderwijsomgeving en is aangeleverd door het Colombiaanse Instituut voor Educatieve Evaluatie. Deze dataset bevat informatie van een cohort van 162.030 Colombiaanse studenten over academische prestaties in het hoger onderwijs, socio-economische status, eerdere academische prestaties, kenmerken van de middelbare school en hun werkstatus.

In hoofdstuk 3 werden systematisch *multilayer perceptron* (MLP's) ontworpen om academische prestaties in het hoger onderwijs te classificeren. De bevindingen toonden aan dat het mogelijk is om academische prestaties nauwkeurig te classificeren als hoog of laag, om de kwaliteit van de classificatie te beoordelen door middel van verschillende kwaliteitsmaatregelen en om de voorspellende bijdrage van de onderzochte voorspellers vast te stellen. In hoofdstuk 4 maken *radial basis function neurale netwerken* (RBFNN's) het mogelijk om academische prestaties op een nauwkeurige en interpreteerbare manier te voorspellen. Met name werd vastgesteld dat eerdere academische prestaties sterk gerelateerd zijn aan academische prestaties in het hoger onderwijs, ongeacht de invloed van SES op eerdere academische prestaties. In hoofdstuk 5 worden *self-organizing maps* (SOM's) gebruikt om studenten te clusteren op basis van de onderzochte voorspellers. Uit de analyse kwamen interessante patronen voort uit eerdere academische resultaten, SES en academische prestaties. Verder werden drie statistisch verschillende groepen studenten geïdentificeerd. Het laatste hoofdstuk van dit proefschrift vat de belangrijkste bevindingen van dit onderzoek samen, beschrijft de belangrijkste bijdragen aan het gebied van de pedagogische wetenschappen en schetst ideeën voor het toekomstig gebruik van artificiële neurale netwerken in onderwijskundig onderzoek.

Thank you…

My doctoral degree is specially dedicated to my nephew Tomás. I would like to thank the following people who have made possible this doctoral dissertation.

First, I would like to express my deepest appreciation to Eva Kyndt and Eduardo Cascallar, my promotors. Eva, thanks for trusting and accepting to be part of this research, even without knowing what the result would be. Thank you for giving me my space to explore, to make mistakes, and to learn from them. As I once told you, you generated a cognitive conflict in me, which allowed me to understand much more of the interesting world of Educational Research. I deeply admire and respect you and your work as a researcher, and when I grow up, I want to be like you.

Eduardo, infinite thanks for allowing me to work with you since 2013. Thank you very much for always playing the devil's advocate, for teaching me that there are no silly questions, for teaching me that I should never assume anything. You are a true teacher, who with his example and dedication knows how to guide his students. It has truly been a privilege to count on you during all this time and I hope one day to positively impact the Educational Sciences, just as you have.

Dad and Mom: there will never be enough words to thank everything you have done for me. Thank you very much for supporting me throughout my life, even if sometimes you have not agreed with some of my decisions. Never forget that I am very proud of you for all that you have accomplished and that I hope that your life lessons will accompany me forever. I love you. Of course, I am also grateful to Andres, my older brother, for his example and support.

Meme, I could write 1000 pages and I would still lack space to thank you for everything you did for me during my Ph.D. Thank you for your comprehension and support, even in the most difficult moments, when I couldn't stand it. But above all, thank you for showing me daily the true meaning of the word love. I love you.

Eva Vermeire, thank you very much for your friendship, for sharing an office with me, and for helping me improve my English. It was a real pleasure to share almost three years with you and to live together all this experience of the doctorate.

Bernarda, I am pleased that you were my partner in crime. You showed me so many things about the culture of Ecuador and you always invited me to take active breaks every day at 4 pm. You are a very special human being and I am sure that you will be able to finish your doctorate in the best way.

Ilke, I want to express my gratitude for all your support, your trust, and your collaboration. You are one of my favorite Belgians. I admire not only who you are as a researcher but also as a person. Never forget that when you want to "get lost" in Colombia, you will be more than welcome.

Katrien, thank you very much for your support and interest. It was always nice to exchange ideas with you and to learn new things. You are a great researcher and I wish you a lot of success in your future career.

Dmitry, it was a pleasure to share beer afternoons and so many nice dinners with you. Thank you very much for your trust and for exchanging so many ideas with me about Educational Research.

Camilo Velez, many thanks for being my academic older brother, for all your support and trust during this time and for being a great example for me.

To my friends from high school, Camilo, Gabriel, and Jorge, your sincere friendship supported me during all this process. It is a blessing to have brothers like you in my life.

Finally, I also thank COLCIENCIAS for having financed this entire research project and the Colombian Institute for Educational Evaluation (ICFES) for having provided the data for this doctoral thesis.

Sincerely yours,

Carlos Felipe

Table of contents

Chapter 1. General Introduction 1

Chapter 2. Socio-economic status and academic performance in higher education: A systematic review 25

Chapter 3. Multistage approach for using artificial neural networks to classify academic performance in higher education 98

Chapter 4. The use of radial basis function neural networks (RBFNNs) to predict academic performance in higher education 149

Chapter 5. A cluster analysis of academic performance in higher education through self-organizing maps (SOMs) 184

Chapter 6. General Discussion 222

Annexes 239

Chapter 1

General Introduction

The understanding of the factors related to academic performance makes it possible to improve the teaching practices, to foster the students' learning process, and to refine the policymaking at different educational levels. Therefore, the analysis of academic performance continues to be an essential research topic within the educational field. A need to better understand the factors related to academic performance gained prominence after the publication of a 1966 report by Coleman and colleagues – sponsored by the National Center for Education Statistics – entitled Equality of Educational Opportunity (EEO). The report concluded that students' socioeconomic characteristics influenced the school performance in the United States, while the variables associated with the school had little or no effect. These results raised great controversy, as many scholars did not understand how the variables related to the school itself would have no impact on academic performance. Thus, many were motivated to conduct studies aimed at corroborating or rather refuting Coleman's and colleagues' hypothesis. According to Gamoran and Long (2007), the stronger influence of family background on academic performance in comparison to the influence of school background is the most cited conclusion of Coleman's report. This fact is illustrated by a recent search on Web of Science, which showed that Coleman's report has been cited over 3600 times since its publication.

The prediction of academic performance has also been of interest in the higher education sector in recent years. Both conceptual and methodological reasons have prompted such interest. To begin with, the relationship between socio-economic status and academic performance has been well established at elementary and secondary levels (e.g., Sirin, 2005). However, this relationship still deserves specific consideration within higher education, given that the participation of low

SES students in the university is increasing. Therefore, to truly respond to the educational needs of low SES students and as a first step in realizing more equity, it seems necessary to unravel the complex nature of how SES and academic performance are related.

It is also the case that the general trend in the literature has been to propose predictive models based on traditional statistical techniques. Although such an approach has been widely accepted among quantitative educational researchers, it is rather uncommon to verify the fulfillment of some of their statistical assumptions, such as homogeneity of variance, normality, and linearity. Therefore, results coming from these traditional predictive models are often methodologically biased. Another limitation of these models based on traditional statistical techniques occurs when vast volumes of information are analyzed. In such cases, the models reduce the reality to a limited set of predictors that are not able to capture the full complexity of the investigated phenomena.

Each of these reasons highlights the need for considering a different methodological approach for the prediction of academic performance in higher education. One powerful example of an alternative way for predictive analysis comes from other disciplines such as business (Chen, Chiang, & Storey, 2012), computer sciences (Hashem et al., 2015), government sector (Kim, Trimi, & Chung, 2014) and management (George, Haas, & Pentland, 2014), where some data mining techniques have been efficiently applied on large datasets. The use of data mining techniques in education research, known as Educational Data Mining (EDM), makes it possible to extract useful information from educational data sets to positively influence the educational field (Romero & Ventura, 2010). Similar to educational sciences in general, predicting academic performance is considered a core topic within the area of EDM (Baker, 2010; Gibson & Ifenthaler, 2017; Romero & Ventura, 2010; Shariri, Husain, & Rashid, 2015). Through the exploration and

discovery of new patterns in large sets of data, EDM facilitates the development of models that can appropriately estimate the present and future states of academic performance (Daniel, 2017; Guo, Zhang, Xu, Shi, & Yang, 2015). Within EDM, artificial neural networks (ANNs) are a well-known technique for the accurate analysis of large educational datasets. ANNs help to develop predictive systems robust to the level of measurement and the statistical distribution of the predictors. Nevertheless, ANNs remain practically unexplored within the educational research field due to conceptual and methodological limitations.

Considering all the above, this doctoral dissertation aims to extend both conceptually and methodologically the use of artificial neural networks as an analysis tool of academic performance in higher education. The expected contribution of this doctoral dissertation is to help to frame the methodology used when applying these machine-learning procedures in educational research. This introduction section presents the theoretical background underlying this research. Next, the objectives, as well as the description of the structure of this doctoral dissertation, are outlined.

Theoretical background

This section begins by describing the use of standardized tests as a measurement of academic performance. Next, several predictors of academic performance in higher education are discussed. Finally, this theoretical background concludes defining ANNs and presenting several of their features when used to analyze academic performance in higher education.

Standardized testing as an indicator of academic performance

Standardized tests are used in different ways, including accountability, appraisal, and assessment (Morris 2011). In particular, Kuncel and Hezlett (2007) suggest that standardized tests assess the level of achievement in subject areas (e.g., mathematics, physics, chemistry) or the

performance in competencies (e.g., verbal, quantitative reasoning, writing). The criticism that standardized testing is often measuring wealth and not academic performance (Mattern, Shawn & Williams, 2008), persists among some educational researchers. However, it could be the case that such criticism only applies to the achievement tests. In this respect, Zwick (2012) has suggested that achievement tests might be indicating the degree to which wealthier students have access to either better information of the test (content hypothesis) or a better preparation for the test (coaching hypothesis).

Two types of performance tests can be distinguished according to the way of reporting and interpreting their outcomes (Zucker, 2003). Norm-referenced testing (NRT) aims to compare the results of a given test-taker to the rest of the test takers (reference group). To make this comparison possible, results are reported as percentile ranks or stanines. Criterion-referenced testing (CRT) indicates the level of performance (e.g., high or low) of each test taker using specific performance standards. Hence, performance tests could be measuring the acquisition of competencies and not necessarily levels of wealth.

The Colombian Institute for Educational Evaluation (*Instituto Colombiano para la Evaluación de la Educación*, ICFES) designs and manages different standardized national tests at various educational levels. For instance, the SABER 5 test is administered at the end of primary school; the SABER 9 test is administered at the end of ninth grade, and the SABER 11 test is taken by all students at the end of their high school. Colombian students must also sit the SABER PRO test when they are about to complete their undergraduate studies. More specifically, the SABER PRO test assesses students' performance in the following areas: critical reading, quantitative reasoning, written communication, civic competences, and English. The higher education sector and the workforce have recognized the importance of students' quantitative reasoning (Roohr,

Graf, & Liu, 2014). Therefore, in this doctoral dissertation, the outcomes in the quantitative reasoning section of the SABER PRO test are chosen as the operationalization of academic performance in higher education.

Prediction of academic performance in higher education

Several theories have been proposed explaining how students grow and change during their university studies (Long, 2012). For instance, Tinto's theory of departure (1993) describes how students' pre-entry characteristics (e.g., their family background, individual attributes, and pre-college schooling) influence their persistence at university. Also, Astin's theory of involvement (1984, 1999) indicates that students' inputs (e.g., socio-economic status, prior academic achievement) interact with environmental elements (e.g., university organizations, peer relations) to define their involvement in the university experience. Starting from the theories mentioned earlier, it can be argued that students begin their university studies with specific characteristics that could explain their adaptation to a new academic setting (De Clercq, Galand, & Frenay, 2017). Moreover, students also have different experiences during their university studies that might have an impact on their academic performance (Astin, 1984; 1999). In line with those theoretical claims, previous research has been conducted to predict students' academic performance from a holistic way, analyzing the contribution of different predictors simultaneously and not by separate (e.g., Zheng, Saunders, Shelley, Mack, & Whalen, 2002; De Clercq et al., 2017). One concluding remark of this body of research is that students' characteristics (Tinto, 1993; Astin, 1984, 1999) such as prior academic achievement, socio-economic status (SES), and high school characteristics; and students' experiences (Astin, 1984, 1999) such as working status during higher education are predictors of academic performance at university. Therefore, this doctoral dissertation will analyze the predictive role of prior academic achievement, socio-economic status (SES), high school

characteristics, and working status on academic performance in higher education. Each of these predictors is discussed in more detail below.

Prior academic achievement

Prior academic achievement is probably the most important predictor of academic performance at university (Li, Chen, & Duanmu, 2010; MacKenzie, & Schweitzer, 2001). In this respect, Warburton, Bugarin, and Nuñez (2001) found that prior academic achievement contributed to reducing the differences in academic performance at university between first-generation students and their peers. Similarly, Crawford (2014) proposed that differences in academic performance at university could be reduced if low SES students improve their prior performance in high school. Furthermore, Marks (2017) has pointed out that SES does not solely explain the influence of prior academic achievement on subsequent students' outcomes at previous stages of students' life. Therefore, it can be argued that prior academic achievement does influence academic performance at university.

Socio-economic status (SES)

Socio-economic status (SES) is perhaps the contextual variable most included in educational research (Sirin, 2005). Education, occupation, and income have been used as indicators for SES in previous educational research (Cowan et al., 2012; Sackett et al., 2009; Van Ewijk & Sleegers, 2010). As such, education, occupation, and income can consistently capture students' socio-economic conditions regardless of the time in which they are observed (Erola, Jalonen, & Lehti, 2016). Furthermore, a moderate effect size between SES and academic performance has been found at elementary and secondary levels (Sirin, 2005; Van Ewijk & Sleegers, 2010). In contrast, the SES-academic performance relationship is less evident in higher

education, as the general trend in the predictive studies has been to include SES as a covariate and not as a predictor. Thus, it seems worthy to unravel the predictive role of SES on academic performance in higher education.

High school characteristics

Several characteristics of high schools may influence students' academic performance at university (Birch & Miller, 2007). For example, high school size, high school-average ability, type of high school (private or public), and the number of students of the high school attending a particular university (Pike & Saupe, 2002). Interestingly, Krashinsky (2014) found that students who attended one less year of high school also had a lower academic performance in university even after controlling for age differences. Besides, Choi, Raley, Muller, and Riegle-Crumb (2008) reported that students whose parents had a higher level of education were more likely to enroll at university, after attending high school with classmates who also had highly educated parents.

Working status

The relationship between working status (i.e., number of working hours) and academic performance has also been discussed in the literature. On the one hand, several authors (Beauchamp, Martineau, & Gagnon, 2016; Rochford, Conolly, & Drennan, 2009; Stinebrickner & Stinebrickner, 2003) have reported a negative relationship between working status and academic performance at university. On the other hand, a positive relationship or no effect at all has also been reported (Nonis & Hudson, 2006; Frischenschlager, Haidinger, & Mitterauer, 2005). Nevertheless, Katsikas (2013) found that family income had a significant influence on the duration of studies for non-working students, but a not significant influence for working students. As the

reasons why students decide to work when attending university are diverse, it can be argued that not only low SES students would work during their studies (Yanbarisova, 2015).

Artificial neural networks (ANNs) and analysis of academic performance

Recent developments in computer science have made it possible to collect and analyze data of increasing size and high dimensionality (Wehrens & Buydens, 2007). Remarkably, those developments have also sparked initial interest in the educational field, where the use of techniques such as artificial neural networks (ANNs) has made it possible to develop accurate and robust predictive systems. Given its crucial role in this doctoral dissertation, it seems necessary to clarify the definition of ANNs as well as some practical implications of their use when analyzing academic performance.

Where do ANNs come from?

Figure 1 depicts that educational data mining (EDM) emerges from integrating three major areas: computer science, education, and statistics; in addition, the intersection of computer science and statistics gives rise to machine-learning, a sub-area related to EDM (Romero & Ventura, 2013). Machine-learning seeks to teach a computer through repeated examples, so it can autonomously solve specific tasks (El Naqa & Murphy, 2015). Artificial neural networks (ANNs) are part of the machine-learning models, and their conceptualization is fully developed below.

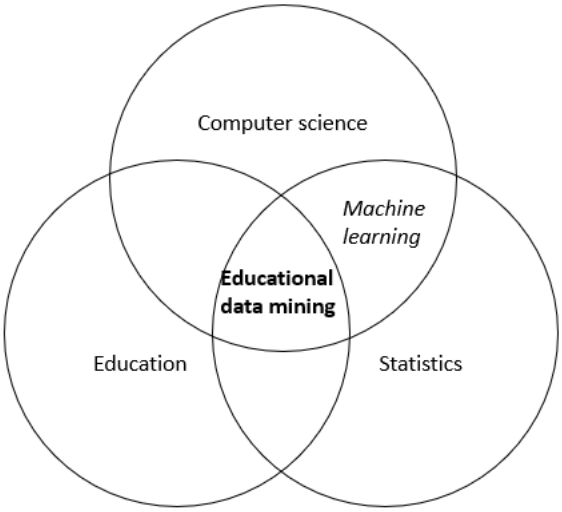


Figure 1. Main areas related to educational data mining (Romero & Ventura, 2013).

What is an artificial neural network?

The general trend in the literature has been to suggest that ANNs resemble both the structure and the functioning of the human brain (e.g., Fausett, 1994; Haykin, 2009; Kriesel, 2007). Hence, it is rather common to find terms such as *neurons* or *synaptic weights* in the definitions of ANNs. Although this biological comparison remains true, the human brain is way more complex than ANNs. Therefore, it seems more straightforward to refer to ANNs in terms of a predictive system as it is described next. Consider the situation of predicting the students' GPA at university using their HSGPA and SES. A first regression model (e.g., multiple linear regression) would allow performing this task and establishing the individual importance of each predictor, as well as the quality of the prediction. In this case, the regression parameters would be obtained using the entire data set at once.

In contrast, a second "regression model" could individually consider each case in the data set and adjust the regression parameters, based on the difference between the predicted and the observed output. Moreover, through this iterative process, the second "regression model" would

identify patterns of information existing in the data set. In other words, the second "regression model" would *learn* from the observed data. An artificial neural network (ANN) is the second "regression model".

Are ANNs different from the traditional statistical techniques?

The easiest way to understand the differences between ANNs and the traditional statistical techniques is realizing the advantages of ANNs for developing predictive systems. First, through ANNs, it is possible to identify all the interactions among the predictor variables to classify and predict outcomes variables, particularly in large data sets (Musso, Kyndt, Cascallar, & Dochy, 2013). Second, ANNs can obtain a complete prediction even when a non-linear relationship exists between independent and dependent variables (Somers & Casal, 2009). Third, predictive systems based on ANNs are robust to the statistical distribution of the data, even when assumptions of linearity, normality, and homogeneity are not fulfilled (Garson, 2014). Therefore, ANNs represent a shift in the traditional paradigm of analyzing academic performance in higher education.

How do ANNs "look like"?

Figure 2 shows a graphical representation of ANNs, indicating three of their defining elements. The first element is the neuron, which is represented with circles in Figure 2. A neuron, or the most fundamental element of any ANN, is simply a mathematical function, which receives an input, transforms it, and returns an output. The second element is the layer of neurons, which appears in Figure 2 as a set of circles of the same color. There are three layers of neurons in ANNs, namely input (which receives the independent variables), hidden (where inputs and outputs are related), and output layer (which returns the dependent variable). The third element is the weights among the neurons, which appear in Figure 2 as the connecting lines between the circles. The

weight is simply the strength of the connection between the neurons as well as the information used by the network to minimize the difference between the predicted and the observed data. In general, any artificial neural network could be characterized by (1) its topology or architecture (i.e., the arrangement of connections between the neurons), (2) its training type (i.e., the approach to obtaining the weights among the neurons), and (3) the mathematical function in each neuron (Fausett, 1994).

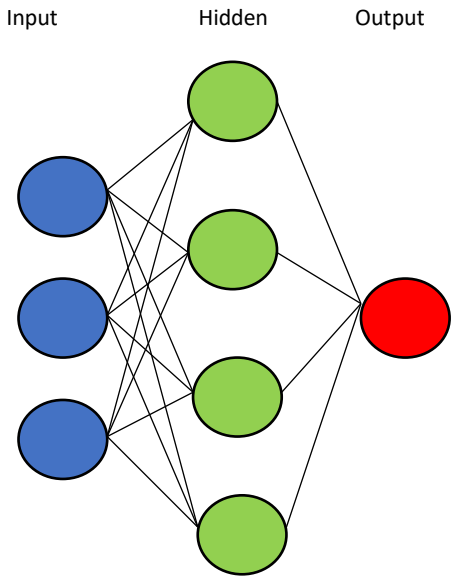


Figure 2. A graphical representation of an artificial neural network.

How can ANNs be trained?

Two distinct approaches can be distinguished to train an artificial neural network (Garson, 2014; Haykin, 2009). Recall that training an artificial neural network is the process of establishing the strengths or weights among its neurons. In the supervised training, the training is achieved by using a set of training vectors, or patterns, each one related to a desired target vector output. The aim of the supervised training is that the network learns from these patterns of information. Therefore, the network adjusts its predictive weights using a learning algorithm (Garson, 2014).

Supervised learning has been used in educational research for predicting students' final grades, test outcomes, and admission to the university (Romero & Ventura, 2010).

Conversely, in the unsupervised training, there is no a priori set of categories into which the patterns must be classified (Garson, 2014). Hence, a sequence of training vectors is provided to the network, but no target vectors are specified. Consequently, the network develops its own representation of the input data (Fausset, 1994). Unsupervised learning has been used in educational research to detect cheating in online assessments (Romero & Ventura, 2010).

How can ANNs be used regarding the analysis of academic performance?

Prediction and clustering are two of the purposes of educational data mining techniques (including ANNs) suggested by Baker (2010). Prediction refers to the estimation of the unknown value of a variable that describes some student's characteristics, trough classification, and regression tasks. In the classification task, the objective is to determine the belonging to specific groups or categories based on the set of predictors. In the regression task, the aim is to find a relationship between a set of predictors and an output variable (Romero & Ventura, 2010). Clustering refers to the task of grouping students from a data set into several subsets. As such, clustering is useful when the most common categories within the data set are unknown (Baker, 2010). The objective of clustering is to create groups according to the specific characteristics of the students in such a way that each group differs from the others (Romero & Ventura, 2010).

This doctoral dissertation

This doctoral dissertation aims to extend both conceptually and methodologically the use of artificial neural networks as an analysis tool of academic performance in higher education. The selected data for analysis is a unique dataset consisting of 162,030 Colombian students and provided by the Colombian Institute for Educational Evaluation (*Instituto Colombiano para la Evaluación de la Educación*, ICFES). Academic performance in higher education is measured as the outcomes in the quantitative reasoning section of the SABER PRO test, while prior academic achievement, socio-economic status (SES), high school characteristics, and working status, are selected as its predictors.

Figure 3 shows how ANNs are used in this doctoral dissertation to analyze academic performance in higher education. According to Figure 3, supervised models are chosen to predict academic performance by carrying out both classification and regression tasks. In addition, unsupervised models are selected to cluster academic performance through a grouping task.

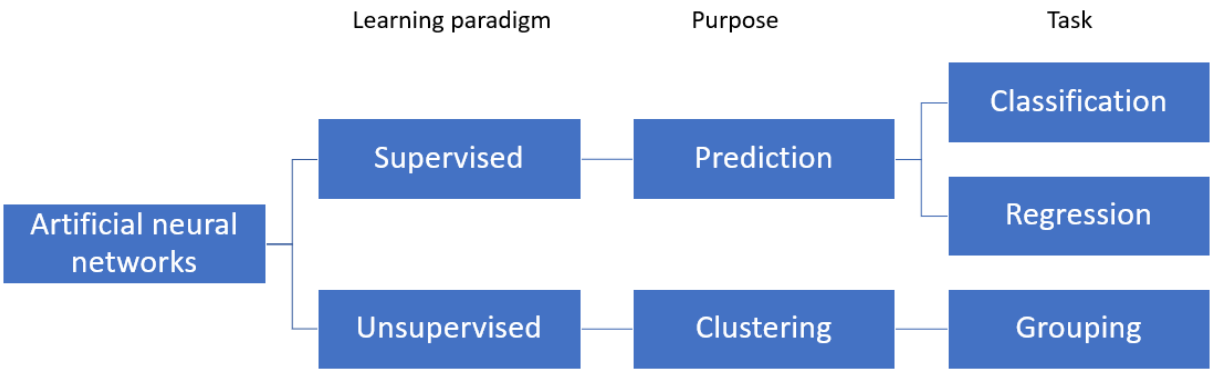


Figure 3. An artificial neural network analysis of academic performance in higher education.

Objectives of this doctoral dissertation

General objective

To extend both the conceptual and methodological understanding of artificial neural networks when used as an analysis tool of academic performance in higher education.

Specific objectives

1. To achieve a better grasp of the existing literature regarding the prediction of academic performance in higher education.
2. To provide a methodological framework for using supervised artificial neural networks to classify academic performance in higher education.
3. To exemplify the use of supervised artificial neural networks to carry out a regression analysis of academic performance in higher education.
4. To identify clusters of academic performance in higher education through unsupervised artificial neural networks.

Structure of this doctoral dissertation

This doctoral dissertation consists of the introduction, four chapters, and a general discussion. In chapter 1, a mixed-methods systematic literature review is carried out for achieving a better grasp of the existing literature regarding the prediction of academic performance in higher education. As such, chapter 2 provided a theoretical cornerstone for developing the predictive models based on ANNs carried out in chapters 3, 4, and 5. Subsequently, each empirical study framed both conceptually and methodologically a different type of artificial neural network. In this way, the empirical studies showed how three specific tasks, namely, classification (chapter 3), prediction (chapter 4), and grouping (chapter 5), can be achieved using ANNs. Lastly, the

discussion section critically integrates the key findings of each of the studies and highlights the contributions of this doctoral dissertation.

Chapter 2. Socio-economic Status and Academic Performance in Higher Education: A Systematic Review. Previous educational research has extensively investigated the relationship between socio-economic status (SES) and academic performance. In higher education, however, this relationship still deserves a comprehensive examination given both conceptual and practical reasons. To attend to this need, a mixed-method systematic literature review of 42 studies is carried out. In the first part, a summative content analysis examines how SES and academic performance are measured. In the second part, a meta-analysis estimates the effect size of the relationship between SES and academic performance in higher education. Findings suggest that SES is measured through education, occupation, income, home resources, and neighborhood resources, while academic performance in higher education is measured through achievement, competencies, and persistence. Furthermore, the meta-analysis reveals a positive yet weak relationship between SES and academic performance in higher education. Prior academic achievement, university experience, and working status are stronger related to academic performance than SES.

Chapter 3. Multistage approach for using artificial neural networks to classify academic performance in higher education. Several limitations have prevented the use of artificial neural networks (ANNs) in educational research. In particular, the absence of standard design rules, the lack of understanding of their quality measures, and the misconception that it is not possible to interpret ANNs models persist among educational scientists. To address such limitations, the present study proposes a multistage approach for the use of ANNs to classify the academic performance of 162,030 university students. Three stages are followed, namely, design,

evaluation, and interpretation. Findings suggest that it is possible to systematically train and test ANNs models to classify students' academic performance as either high-performance (accuracy of 82%) or low-performance (accuracy of 75%); to assess the quality of the derived predictions using several quality measures; and to provide clear and transparent interpretations of the predictive models based on ANNs. Furthermore, prior academic achievement, socio-economic conditions, and high-school characteristics continue to be weighty predictors of students' academic performance in higher education.

Chapter 4. The use of radial basis function neural network (RBFNNs) to predict academic performance in higher education. While educational researchers know regression analysis widely, its use still copes with several important limitations. In particular, the difficulty in interpreting regression coefficients derived from categorical predictors, as well as the undesired effects of violating statistical assumptions, highlight the need for exploring a different prediction approach. This methodological study shows the potential of the radial basis function neural network (RBFNNs) as a predictive tool for educational data. As an illustration case, RBFNNs is used to predict the academic performance in a cohort of 162,030 Colombian university students based on their prior academic achievement, socio-economic status (SES), high-school characteristics, and working status. Findings suggest that RBFNNs are easy to design and capable of satisfactorily represent and interpret the relationship between academic performance and its predictors, regardless of their level of measurement and statistical distribution. As such, RBFNNs provides a promising alternative to carry out predictions of educational data.

Chapter 5. A cluster analysis of academic performance in higher education through self-organizing maps (SOMs). Cluster analysis is a powerful tool to explore educational data in order to create groups of students, teachers, or institutions according to their similarities. Self-organizing

maps (SOMs) is one of the existing clustering methods, which makes it possible to cluster large data sets, to identify hidden patterns in the data, and to visualize complex relationships among the investigated variables. Interestingly, SOMs find relationships in data sets where information on the output is unknown. Therefore, it seems worthy to further extend the use of SOMs to cluster educational data. This study applies self-organizing maps (SOMs) to a cohort of 162,030 university students to determine clusters based on their prior academic achievement, socio-economic status (SES), and academic performance in higher education. Findings suggest that interesting patterns exist between prior academic achievement, academic performance in higher education, and SES. Starting from these variables, students can be clustered into three groups, which are statistically different among them.

Chapter 6. General discussion. In this final section, the seven key findings of this doctoral dissertation are listed and discussed. Next, conclusions are implications for practice as well as the limitations of this work are highlighted. Finally, the suggestions for future research are given.

References

Astin, A. W. (1984). Student involvement: A developmental theory for higher education. *Journal of College Student Personnel, 25*, 297-308.

Astin, A. W. (1999). Student involvement: A developmental theory for higher education. *Journal of College Student Development, 40,* 518-529.

Baker, R. S. J. D. (2010). Data mining for education. In McGaw, B., Peterson, P., Baker, E. (Eds.) *International Encyclopedia of Education*: *Vol.7* (pp.112-118). Oxford, UK: Elsevier.

Beauchamp, G., Martineau, M., & Gagnon, A. (2016). Examining the link between adult attachment style, employment and academic achievement in first semester higher education. *Social Psychology of Education, 19,* 367-384. doi:10.1007/s11218-015-9329-3

Birch, E. R., & Miller, P. W. (2006). Student outcomes at university in Australia: A quantile regression approach. *Australian Economic Papers*, *45*, 1-17. doi:10.1111/j.1467-8454.2006.00274.x

Chen, H., Chiang, R. H., & Storey, V. C. (2012). Business intelligence and analytics: From big data to big impact. *MIS Quarterly, 36*, 1165-1188.

Choi, K. H., Raley, R. K., Muller, C., & Riegle-Crumb, C. (2008). A longitudinal analysis of contextual exposure to peers with college-educated parents and students' college enrollment. *Social Science Quarterly, 89*, 846-866. doi:10.1111/j.1540-6237.2008.00587.x

Coleman, J., Campbell, E., Hobson, C., McPartland, F., Mood, A., & Weinfeld, F. (1966). *Equality of educational opportunity*. Washington DC: U.S. Government.

Cowan, C. D., Hauser, R., Kominski, R., Levin, H., Lucas, S., Morgan, S., Spencer, M.B., & Chapman, C. (2012). *Improving the measurement of socioeconomic status for the national assessment of educational progress: A theoretical foundation*. Retrieved from National Center for Education Statistics website https://nces.ed.gov/nationsreportcard/pdf/researchcenter/Socioeconomic_Factors.pdf

Crawford, C. (2014). *Socio-economic differences in university outcomes in the UK: Drop-out, degree completion, and degree class* (Institute for Fiscal Studies Working Paper No. W14/31). Retrieved from Institute for Fiscal Studies website https://www.ifs.org.uk/uploads/publications/wps/WP201431.pdf

Daniel, B. K. (2017). *Big data and learning analytics in higher education*. Switzerland: Springer International Publishing. doi: 10.1007/978-3-319-06520-5

De Clercq, M., Galand, B., & Frenay, M. (2017). Transition from high school to university: a person-centered approach to academic achievement. *European Journal of Psychology of Education*, *32*, 39-59. doi: 10.1007/s10212-016-0298-5

El Naqa, I., & Murphy, M. J. (2015). What is machine learning? In I. El Naqa, R. Li, & M.J. Murphy (Eds.) *Machine learning in radiation oncology* (pp. 3-11). doi: 10.1007/978-3-319-18305-3

Erola, J., Jalonen, S., & Lehti, H. (2016). Parental education, class, and income over early life course and children's achievement. *Research in Social Stratification and Mobility, 44*, 33-43. doi:10.1016/j.rssm.2016.01.003

Fausett, L. V. (1994). *Fundamentals of neural networks*. New Jersey, USA: Prentice-Hall.

Frischenschlager, O., Haidinger, G., & Mitterauer, L. (2005). Factors associated with academic success at Vienna Medical School: Prospective survey. *Croatian Medical Journal, 46,* 58-65.

Gamoran, A., & Long, D. A. (2007). Equality of educational opportunity. A 40-year retrospective. In R. Teese, S. Lamb, & M. Duru-Bellat (Eds.), *International studies in educational inequality, theory, and policy* (pp. 23-47). Dordrecht, The Netherlands: Springer. doi:10.1007/978-1-4020-5916-2_2

Garson, G. D. (2014). *Neural network models*. Asheboro, NC: Statistical Associates Publishers.

George, G., Haas, M. R., & Pentland, A. (2014). Big data and management. *Academy of Management Journal*, *57,* 321-326. doi:10.5465/amj.2014.4002

Gibson, D., & Ifenthaler, D. (2017). Preparing the next generation of education researchers for big data in higher education. In Daniel B.K. (Eds), *Big data and learning analytics in higher education* (pp. 29-43). Switzerland: Springer International Publishing, doi: 10.1007/978-3-319-06520-5

Guo, B., Zhang, R., Xu.G., Shi, C., & Yang. L. (2015). Predicting students' performance in educational data mining. *Proceedings of the 2015 International Symposium on Educational Technology*. International Symposium on Educational Technology. Wuhan, 27-29 July, 125-128, doi: 10.1109/ISET.2015.33

Hashem, I. A. T., Yaqoob, I., Anuar, N. B., Mokhtar, S., Gani, A., & Khan, S. U. (2015). The rise of "big data" on cloud computing: Review and open research issues. *Information Systems*, *47*, 98-115, doi: 10.1016/j.is.2014.07.006

Haykin, S. S. (2009). *Neural networks and learning machines* (Vol. 3). Upper Saddle River, NJ, USA: Pearson.

Katsikas, E. (2013). The impact of students' working status on academic progress: Assessing the implications of policy change in Greece. *Journal of Education and Work, 26,* 539-569. doi:10.1080/13639080.2012.691606

Kim, G. H., Trimi, S., & Chung, J. H. (2014). Big-data applications in the government sector. *Communications of the ACM, 57,* 78-85, doi:10.1145/2500873

Krashinsky, H. (2014). How would one extra year of high school affect academic performance in university? Evidence from an educational policy change. *Canadian Journal of Economics/Revue Canadienne d'économique, 47,* 70-97. doi:10.1111/caje.12066

Kriesel, D. (2005). *A brief introduction to neural networks*. Retrieved from: http://www.dkriesel.com/_media/science/neuronalenetze-en-zeta2-2col-dkrieselcom.pdf

Kuncel, N. R., & Hezlett, S. A. (2007). Standardized tests predict graduate students' success. *Science, 315,* 1080-1081, doi: 10.1126/science.1136618

Li, G., Chen, W., & Duanmu, J. L. (2010). Determinants of international students' academic performance: A comparison between Chinese and other international students. *Journal of Studies in International Education, 14,* 389-405. doi:10.1177/1028315309331490

Long, D. (2012). Theories and models of student development. In L. J. Hinchliffe & M. A. Wong (Eds.), *Environments for student growth and development: Librarians and student affairs in collaboration* (pp. 41-55). Chicago: Association of College & Research Libraries.

Marks, G. N. (2017). Is SES really that important for educational outcomes in Australia? A review and some recent evidence. *The Australian Educational Researcher, 44,* 191-211. doi: 10.1007/s13384-016-0219-2.

Mattern, K., Shawn, E., & Williams, F. (2008). Examining the relationship between the SAT®, high school measures of academic performance, and socioeconomic status: Turning our attention to the unit of analysis. *College Board Research Notes, 36,* 1-10.

McKenzie, K., & Schweitzer, R. (2001). Who succeeds at university? Factors predicting academic performance in first-year Australian university students. *Higher Education Research and Development, 20,* 21-33. doi:10.1080/07924360120043621

Morris, A. (2011). Student standardized testing: current practices in OECD countries and a literature review. *OECD Education Working Papers, 65,* 1-51.

Musso, M. F., Kyndt, E., Cascallar, E. C. & Dochy F. (2013) Predicting general academic performance and identifying the differential contribution of participating variables using artificial neural networks. *Frontline Learning Research, 1,* 42-71 doi:10.14786/flr.v1i1.13.

Nonis, S. A., & Hudson, G. I. (2006). Academic performance of college students: Influence of time spent studying and working. *Journal of Education for Business, 81*, 151-159. doi:10.3200/joeb.81.3.151-159

Pike, G. R., & Saupe, J. L. (2002). Does high school matter? An analysis of three methods of predicting first-year grades. *Research in Higher Education, 43*, 187-207.

Rochford, C., Connolly, M., & Drennan, J. (2009). Paid part-time employment and academic performance of undergraduate nursing students. *Nurse Education Today, 29*, 601-606. doi:10.1016/j.nedt.2009.01.004

Romero, C., & Ventura, S. (2010). Educational data mining: review of the state of the art. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 40,* 601-618, doi: 10.1109/TSMCC.2010.2053532

Romero, C., & Ventura, S. (2013). Data mining in education. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 3,* 12-27. doi: 10.1002/widm.1075

Roohr, K. C., Graf, E. A., & Liu, O. L. (2014). Assessing quantitative literacy in higher education: An overview of existing research and assessments with recommendations for next-generation assessment. *ETS Research Report Series, 2,* 1-26

Sackett, P. R., Kuncel, N. R., Arneson, J. J., Cooper, S. R., & Waters, S. D. (2009). Does socioeconomic status explain the relationship between admissions tests and postsecondary academic performance? *Psychological Bulletin, 135,* 1-22. doi:10.1037/a0013978

Shahiri, A. M., Husain, W., & Rashid, N.A. (2015). A review on predicting student's performance using data mining techniques. *Procedia Computer Science, 72,* 414-422, doi:10.1016/j.procs.2015.12.157

Sirin, S. (2005). Socioeconomic status and academic achievement: A meta-analytic review of research. *Review of Educational Research, 75,* 417-453. doi:10.3102/00346543075003417

Somers, M. J., & Casal, J. C. (2009). Using artificial neural networks to model nonlinearity: The case of the job satisfaction job performance relationship. *Organizational Research Methods, 12,* 403-417.

Stinebrickner, R., & Stinebrickner, T. R. (2003). Working during school and academic performance. *Journal of Labor Economics, 21,* 473-491. doi:10.1086/345565

Tinto, V. (1993). *Leaving college. Rethinking the causes and cures of student attrition.* Chicago, IL: The University of Chicago Press.

Van Ewijk, R., & Sleegers, P. (2010). The effect of peer socioeconomic status on student achievement: A meta-analysis. *Educational Research Review, 5,* 134-150. doi:10.1016/j.edurev.2010.02.001

Warburton, E., Bugarin, R., & Nunez, A. (2001). *Bridging the gap: Academic preparation and postsecondary success of first-generation students* (NCES 2001–153). Washington, DC: National Center for Education Statistics, U.S. Government Printing Office.

Wehrens, R., & Buydens, L. M. (2007). Self-and super-organizing maps in R: the Kohonen package. *Journal of Statistical Software, 21,* 1-19. doi: 10.18637/jss.v021.i05

Yanbarisova, D. M. (2015). The effects of student employment on academic performance in Tatarstan higher education institutions. *Russian Education & Society, 57,* 459-482. doi:10.1080/10609393.2015.1096138

Zheng, J. L., Saunders, K. P., Shelley, I. I., Mack, C., & Whalen, D. F. (2002). Predictors of academic success for freshmen residence hall students. *Journal of College Student Development, 43*, 267-283.

Zucker, S. (2003). Fundamentals of standardized testing. Retrieved from Pearson Education Website. http://images.pearsonassessments.com/images/tmrs/tmrs_rg/Fundamentals_of_Standardized_Testing.pdf

Zwick, R. (2012). The role of admissions test scores, socioeconomic status, and high school grades in predicting college achievement. *Pensamiento Educativo. Revista de Investigación Educacional Latinoamericana, 2,* 23-30. doi:10.7764/pel.49.2.2012.3

Chapter 2

Socio-economic status and academic performance in higher education: A systematic review[1]

Abstract

Previous educational research has extensively investigated the relationship between socio-economic status (SES) and academic performance. In higher education, however, this relationship still deserves a comprehensive examination given both practical and conceptual reasons. To attend to this need, a mixed-methods systematic literature review of 42 studies has been carried out. In the first part, a summative content analysis examines how SES and academic performance are measured. In the second part, a meta-analysis estimates the effect size of the relationship between SES and academic performance in higher education. Findings suggest that SES is measured through education, occupation, income, household resources, and neighborhood resources, while academic performance in higher education is measured through achievement, competencies, and persistence. Furthermore, the meta-analysis reveals a positive yet weak relationship between SES and academic performance in higher education. Prior academic achievement, university experience, and working status are more strongly related to academic performance than SES.

Keywords: Socio-economic status; Academic performance; Higher education; Meta-analysis

[1] The text in this chapter has been published as Rodríguez-Hernández C.F., Cascallar, E., Kyndt, E. (2020). Socio-economic status and academic performance in higher education: A systematic review, *Educational Research Revie*w, *29*, 100305. doi: 10.1016/j.edurev.2019.100305

1. Introduction

Over the past years, the student population applying to and entering university has become more diverse in terms of social, cultural and economic capital, age, nationality (Morlaix & Suchaut, 2014), prior education, and academic achievement (Anderton, Evans, & Chivers, 2016). Moreover, in many countries, social changes have also contributed to changes in higher education systems; thus, although there is still a long way to go, participation from students from low social and economic backgrounds in higher education is increasing (Hansen & Mastekaasa, 2006). In order to achieve a better understanding of these changes in higher education systems, several researchers have explored the relationship between socio-economic status (SES) and academic performance, finding a weak to moderate relationship (e.g., Richardson, Abraham, & Bond, 2012; Sackett et al., 2012; Westrick, Le, Robbins, Radunzel, & Schmidt, 2015).

Given the increasing participation of students from low SES backgrounds and the weak to moderate relation between SES and academic performance, one could wonder whether we still need to study the relationship between SES and academic performance in higher education. However, both practical and conceptual reasons exist that warrant continued attention. Firstly, the common trend in the literature has been to use SES as a covariate instead of establishing, in a more comprehensive way, its influence on students' experiences and outcomes (McKenzie & Schweitzer, 2001; Walpole, 2003). By doing the latter, the higher education sector can gain a deeper understanding on how SES and academic performance are related, which might be necessary in order to deal appropriately with the increased diversity in the student population. Secondly, recent meta-analyses in higher education (Richardson et al., 2012; Schneider & Preckel, 2017; Westrick et al., 2015) have focused primarily on calculating the effect size of the relationship between SES and academic performance, possibly ignoring the influence of several other students'

and institutional characteristics on this relationship. Thirdly, based on analytic schemes such as the Astin's input- environment-output model (Astin, 1993), it could be argued that the relationship between SES and academic performance in higher education changes when additional variables are taken into account. As such, it seems inadequate to explore students' academic performance as a single-factor phenomenon (De Clercq, Galand, Dupont, & Frenay, 2013).

Although studies reviewing the relationship between SES and academic performance in primary and secondary schools are available (e.g., Sirin, 2005), it cannot be assumed that these results are generalizable to the context of higher education. By carrying out a mixed-methods systematic literature review, this study seeks to fill the gap that exists in the extensive literature devoted to examining the relationship between SES and academic performance in higher education. This article begins by presenting several issues pertaining to the definition of SES and academic performance, and then describes previous research in the literature regarding their relationship. Next, the methodology of this study is outlined in detail, explaining how both summative content analysis and meta- analysis are carried out. Subsequently, the answers to the research questions addressed in this research are presented. The article concludes with the discussion and implications of the key findings of this systematic literature review.

2. Theoretical framework

2.1. Conceptualizing and measuring socio-economic status (SES)

The understanding of students' socio-economic conditions became a major concern for educational researchers when low academic performance at school was observed in students whose parents had low income, low levels of education, and were employed at low-status jobs (Cowan et al., 2012). Although SES can be considered as one of the most commonly used variables in educational research (Sirin, 2005), it has been conceptualized in different ways in the literature.

For instance, Chapin (as cited in White, 1982) defined SES in 1928 as: “the position that an individual or family occupies with reference to the prevailing average of standards of cultural possessions, effective income, material possessions, and participation in group activity in the community” (p. 99). Mueller and Parcel (1981) defined SES as the position of an individual, family, or group on a hierarchy based on economic, power, and prestige dimensions. More recently, SES has been defined as the amount of economic, social, and cultural resources available to one student (Cowan et al., 2012; De Clercq, Galand, & Frenay, 2017).

The different dimensions of SES have been operationalized using either single indicators, multiple indicators analyzed separately, or several indicators combined in a composite score (Australian Bureau of Statistics, 2011; Cowan et al., 2012; Shavers, 2007). Moreover, the indicators of SES can be observed at several levels, namely, the individual, family, or area levels (Australian Bureau of Statistics, 2011; Krieger, Williams, & Moss, 1997). At the individual level, education, occupation, and income have been used as indicators for SES in previous educational research (Cowan et al., 2012; Sackett, Kuncel, Arneson, Cooper, & Waters, 2009; Van Ewijk & Sleegers, 2010). Education, occupation, and income can consistently capture students' socio-economic conditions regardless of the time in which they are observed (Erola, Jalonen, & Lehti, 2016). In addition, these measurements are easy to interpret and communicate (Cowan et al., 2012). At the family level, household resources have been suggested as the fourth indicator for SES (Sirin, 2005). Household resources refer to possessions such as cars, books, computers, and musical instruments (De Clercq et al., 2017; Pedrosa, Dachs, Maia, Andrade, & Carvalho, 2007). Finally, at the area level, neighborhood resources have been reported as the fifth indicator for SES (Australian Bureau of Statistics, 2011; Cowan et al., 2012; Shavers, 2007). Interestingly, financial and social resources that do not come exclusively from the family can also be related to students’

academic performance (Cowan et al., 2012). Such is the case, for example, with neighborhood characteristics and resources like the degree of urbanization (Hansen & Mastekaasa, 2006), and the number of parks and libraries in the area where students live (Cowan et al., 2012).

How education, occupation, income, and household resources interact in the measurement of SES, however, is rather complex. In particular, education has been the most commonly used indicator to assess SES (Australian Bureau of Statistics, 2011; Shavers, 2007) because of its relationship with other aspects of socio-economic status (Erola et al., 2016; Galobardes, Shaw, Lawlor, Lynch, & Smith, 2006). In fact, higher levels of education are related to the subsequent benefits they offer for a person's life and wellbeing, such as a better job, working conditions, and higher-income (Shavers, 2007). Similarly, occupation is also commonly used as an indicator of SES mainly due to its relationship with education and income (Erola et al., 2016; Ganzeboom, De Graaf, & Treiman, 1992). In this case, income represents the amount of social and economic resources a student can have (Australian Bureau of Statistics, 2011; Galobardes et al., 2006; Sirin, 2005), whereas household resources, therefore, can also indicate whether a student's home situation is adequate for learning (Van Ewijk & Sleegers, 2010).

In summary, SES is a broad concept that encompasses two primary dimensions: prestige and resources (Krieger et al., 1997). The first dimension determines the hierarchical position of an individual in a society (Mueller & Parcel, 1981), while the second dimension determines the economic, social, and cultural resources which an individual has access to (Cowan et al., 2012; De Clercq et al., 2017). In addition, education, occupation, income, and household resources have been widely used as measurements to assess SES.

2.2.Measuring academic performance in higher education

When searching for the definition of students' academic performance in higher education, the lack of consensus in the educational literature is evident. Terms such as performance, achievement, and success are used interchangeably among educational researchers without any specific reason (e.g., Casillas et al., 2012; Rochford, Connolly, and Drennan, 2009; Tracey, Allen, & Robbins, 2012). Also, an operationalization (e.g., Grade Point Average) rather than a conceptual definition is mostly reported when defining academic performance in higher education. Despite this lack of consensus, academic achievement, competencies, and persistence have been used as separate, although interrelated, measurements to assess students' academic performance in higher education. Simply looking at academic achievement does not necessarily encompass or represent students' acquisition of competencies nor their persistence (York, Gibson, & Rankin, 2015).

Academic achievement can be defined as the attainment of either medium- or long-term educational goals (Yusuf, 2002). In this respect, Li, Chen, and Duanmu (2010) have pointed out that prior academic achievement is strongly related to students' academic performance at university. As a matter of fact, a considerable number of studies have reported the explanatory role of prior academic achievement in academic performance at university (e.g. Betts, Elder, Hartley, & Blurton, 2008; Byrne & Flood, 2008; Casillas et al., 2012; McKenzie & Schweitzer, 2001; Pike & Saupe, 2002; Roberts, 2007). Furthermore, a competency is a "performance capacity to do as well as to know which is judged by some level or standard of performance" (Shavelson, 2010, p. 44). In particular, higher education aims at developing both specific and generic competencies (Sadler, 2013). Undoubtedly, a deeper understanding of academic performance in higher education requires the assessment of both generic and specific student competencies (Blömeke, Zlatkin-Troitschanskaia, Kuhn, & Fege, 2013). Consequently, the assessment of competencies has arisen

in many countries at different stages of the higher education learning process (Zlatkin-Troitschanskaia, Shavelson, & Kuhn, 2015). Finally, Tinto's (1993) theory of departure indicates that students persist when they are integrated into both the academic and social systems of the university. Persistence can be understood as the students' academic progression towards degree completion regardless of institutional transfers, academic programs, or institutional contexts (York et al., 2015). The dropout rate has usually been suggested as an indicator of persistence in higher education (Hilton, 1982; Tinto, 1975, 1993).

2.3.The relationship between SES and academic performance

One of the most crucial turning points in educational research during the 20th century was the publication of a 1966 report by Coleman and colleagues entitled Equality of Educational Opportunity (EEO). The report suggested that high school characteristics were unrelated to academic performance in the USA, but students' socio-economic conditions were. Consequently, many educational researchers have carried out studies aimed at understanding Coleman's main findings.

On the one hand, several authors have postulated theoretical frameworks, such as Astin's (1984,1999) student involvement theory or Bourdieu's (1986) social capital theory, which could explain how SES and academic performance are related. Regarding social capital theory (Bourdieu, 1986), Dika and Singh (2002) critically reviewed the literature (published between 1986 and 2001) that relates social capital to educational outcomes and identified several problems with the conceptualization and measurement of social capital as a predictor of academic performance. In particular, sources of social capital are often confused with the resources and opportunities coming from it; thus, there is no clear distinction between possession and activation of social capital (Dika & Singh, 2002). Additionally, the selection of cross-sectional data has made

it difficult to determine how social resources and educational outcomes are related (Dika & Singh, 2002). Therefore, it cannot be entirely accepted that social capital explains how SES is related to academic performance (Jæger, 2011; Sullivan, 2001).

On the other hand, several meta-analytic studies have been conducted in different educational settings. Focusing on the elementary and secondary levels, the main objectives of these meta-analyses (e.g., Sirin, 2005; Van Ewijk & Sleegers, 2010; White, 1982) have been: (1) to determine the effect size of the relationship between SES and academic performance; and (2) to identify which factors could moderate the relationship between SES and academic performance. Regarding the first objective, the findings show a positive, albeit moderate, relationship, as indicated by average correlations of .343 (S.D. = .204; White, 1982), .299 (S.D. = .169; Sirin, 2005) and .32 (S.E. = .016; Van Ewijk & Sleegers, 2010). With respect to the second objective, the results suggest that methodological factors such as the unit of analysis (Sirin, 2005; White, 1982), the definition of SES (White, 1982), the source of the SES data, the range of the SES, and the type of SES-performance measure (Sirin, 2005), moderate the relationship between SES and academic performance.

Delving into higher education, one of the objectives of these meta-analyses (e.g., Richardson et al., 2012; Schneider & Preckel, 2017; Westrick et al., 2015) has been to explore the effect size of the relationship between SES and academic performance. Richardson et al. (2012) found a small correlation between SES and academic performance ($r = .11$, 95% CI [.08, .15]). Similarly, Westrick et al. (2015) reported that SES is weakly related to first-year GPA ($r = .24$, 95% CI [.24, .25]) and second-year retention at university ($r = .10$, 95% CI [.09, .11]). Finally, SES was ranked in the 68th place among 105 variables associated with academic performance in higher education (Schneider & Preckel, 2017). A concluding remark of this body of research is

that there is a weak to moderate relationship between SES and academic performance in higher education. However, a more comprehensive exploration of this relationship is still missing in the educational literature.

2.4.Mediators of the relationship between SES and academic performance

Several theories have been proposed to explain how students grow and change during their university studies (Long, 2012). A well-known example within this body of literature is Astin's (1984, 1999) theory of involvement. Students' involvement refers to the extent to which a student invests energy into his or her university experience —the greater the student's involvement, the greater the students' learning and development (Astin, 1984, 1999). One analytic scheme drawn from Astin's theory of involvement is the input-environment-output (IEO) model (Astin, 1993). The I-E-O model suggests that students' educational outcomes are defined by the students' inputs (e.g., their demographic characteristics, prior academic achievement), the environmental elements (e.g., university organizations, peer relations), and the interaction among students' inputs and environmental elements (Astin, 1993). Most of the predictive studies in higher education lack of a theoretical foundation, being more empirically-based than theory-driven. Nevertheless, the I-E-O model seems to be a suitable approach to analyze the literature in higher education regarding the prediction of academic performance for two reasons. First, the I-E-O model allows for the investigation of the direct and indirect influence (via environmental characteristics) of students' inputs on students' educational outcomes. Second, the I-E-O model recognizes the longitudinal nature of the student retention process and provides a framework for the investigation of it (Kelly, 1996).

Starting from the I-E-O model, predictive studies in higher education can be classified into two types. The first type of predictive studies, which is most likely the most frequent one in higher

education, has examined academic performance as an "input-output" process, a common approach in the educational field. A review of the literature on the input-output analysis of the schools was carried out by Glasman and Biniaminov (1981). The authors concluded that inputs can be categorized into either student or school type, while the outputs can be classified as either cognitive or non-cognitive. Furthermore, a causal model including both the direct and indirect effects of inputs on outputs was also proposed by Glasman and Biniaminov (1981). The second type of predictive studies in higher education has investigated academic performance as an "input-environment-output" process. A similar analytic scheme at the school level is the context, input, process, and output model proposed by Scheerens (1990). In brief, context here refers to the school environment as well as the policy measures at a higher administrative level, while input relates to the available resources, teacher qualifications, and student characteristics. Process includes curriculum, school organization, and school climate, while output is generally defined in terms of students' achievement.

Furthermore, the I-E-O model can frame the analysis of the mediators of the relationship between SES and academic performance in higher education. First, several "input-output" studies in higher education (e.g., Crawford, 2014; Stratton & Wetzel, 2011; Warburton, Bugarin, & Nuñez, 2001) have shown that (1) prior academic achievement is strongly related to academic performance in university and (2) prior academic achievement might diminish the strength of the SES-academic performance relationship. Although it has been amply documented that SES determines prior academic achievement, Marks (2017) has suggested that the influence of prior academic achievement on students' outcomes is not solely explained by their SES at previous stages of life. Therefore, the first mediator investigated in this systematic literature review is prior academic achievement.

Second, the "input-process-output" studies in higher education have suggested that university experience (Gerken & Volkwein, 2000; Smith, 2016; Walpole, 2003) does influence the relationship between SES and academic performance. University experience refers to how a student connects to the academic environment of the university (Astin, 1999). More specifically, perception of the learning environment, peer support, and institutional commitment could define university experience (Astin, 1999; De Clercq et al., 2013). Thus, the second mediator investigated in this systematic literature review is university experience.

Third, although Astin (1999) suggested that holding a part-time job on campus could have a beneficial influence on students' retention, the explanatory role of working status has usually been investigated separately from the university experience. In this respect, the general trend in the literature has been to analyze the influence of worked hours on students' academic performance (e.g., Nonis & Hudson, 2006; Rochford et al., 2009; Stinebrickner & Stinebrickner, 2003). Nevertheless, the reasons why students decide to work when attending university are diverse; so, it can be argued that not only would low SES students work during their studies (Yanbarisova, 2015). Hence, the third mediator investigated in this systematic literature review is working status.

3. Present study

Past educational research has contributed to the understanding of the relationship between SES and academic performance. However, recent meta-analyses in higher education (Richardson et al., 2012; Schneider & Preckel, 2017; Westrick et al., 2015) are limited, for two main reasons. Firstly, both SES and academic performance had been operationalized in these studies using only one indicator, which narrowed the understanding regarding how additional measures to assess such complex terms are related. Secondly, these meta-analyses did not explore how SES is first related to several student characteristics and, subsequently, is related to academic performance. This

systematic review focuses on investigating the relationship between socio-economic status and academic performance in higher education in a more comprehensive manner. The first objective of this study is to analyze the different measures of SES and academic performance in higher education. Thus, the first research question addressed in this study is: *How are SES and academic performance in higher education measured?* The second objective of this study is to determine the mediating role of several factors on the relationship between SES and academic performance in higher education. Therefore, the second pair of research questions of this study are: *(a) What is the relationship between SES and academic performance in higher education? And (b) Is the relationship between SES and academic performance in higher education mediated by (i) prior academic achievement, (ii) university experience, and/or (iii) working status?* To answer these research questions, a mixed-methods research synthesis (Heyvaert, Maes, & Onghena, 2013) of the selected studies has been carried out. A mixed-methods approach makes it possible to integrate both qualitative analyses (e.g., for this study, summative content analysis) and quantitative analyses (e.g., for this study, meta-analysis) of the results of the studies in order to obtain conclusions about the current state of the art of the literature (Heyvaert et al., 2013).

4. Method

This systematic review proceeded in three phases. In the first phase, an extensive literature search within several scientific databases was conducted. Next, the relevant literature retrieved from these databases was selected for inclusion according to several criteria. In the second phase, the quality of the selected studies was critically appraised, and finally, the primary studies were analyzed following the guidelines of Aveyard (2014) and the performance of both a summative content analysis and meta-analysis.

4.1.Literature search and literature selection

In this systematic review, four databases were consulted: ERIC, Scopus, SSCI, and PsycArticles, using combinations of the following search terms: “higher education”, “academic performance”, “academic achievement”, and “academic outcomes”. The initial search yielded 28,414 non-unique studies, as shown in Table 1. The selection of the literature was based on eight criteria for inclusion. As reported by Gamoran and Long (2007), the citations of Coleman's report increased again by the end of the ’90s and achieved an average of 55 citations per year from 2000, which represents a new interest for studying academic performance since then. Thus, studies published after 2000 (Criterion 1) were considered for subsequent analysis. Second, only empirical studies exploring academic performance in higher education were included in this analysis (Criterion 2). Studies focusing on specific subgroups of students (i.e., students with disabilities; online learners) were excluded (Criterion 3). As this systematic literature review focused on variables related to SES, studies concerning learning styles were not included (Criterion 4). Similarly, studies regarding age differences (Criterion 5), gender differences (Criterion 6), and ethnic differences (Criterion 7) were not included in this analysis. Finally, only studies that explicitly report the relationship between SES and academic performance were selected for further analysis (Criterion 8).

The selection process of the initial 28,414 studies was carried out in six steps. In step 1, duplicate studies were eliminated using the EndNote software, leaving 24,246 studies. In step 2, studies published before 2000 were excluded, leaving 9,928 studies. Next, in step 3, studies without a date, a wrong date, or those that were not written in English or Spanish were eliminated, leaving 9,658 studies. As expected, a summative content analysis requires a full understanding of the reviewed studies. Therefore, in step 4, the title and the abstract of the remaining studies were

screened using the first seven criteria for inclusion presented above. As a result of this fourth step, 208 studies were retained for further analysis. In step 5, a total of 202 studies were scanned/read diagonally as six full texts articles could not be retrieved. This fifth step, using the same inclusion criteria, led to a further reduction in the number of studies to a total of 100. In the final step, the full texts of these 100 studies were read in- depth, and 69 additional studies were excluded using Criterion 8 presented above. As a result of the selection process, 31 studies remained. As any systematic review is a challenging and time-consuming task, 18 months passed between the first literature search and the writing of the first draft of this article. Hence, an update of the literature search was carried out on May 2017. In total, 1,652 new hits were retrieved, and after using the previously mentioned selection criteria, six additional studies were selected. Finally, a back-tracing process of the 37 resulting studies was performed to identify and select additional relevant studies. The outcome of this process was the selection of ten further studies, and as a result of the selection process, a total of 47 primary studies were considered for critical appraisal.

4.2.Critical appraisal

The quality of the selected primary studies (n = 47) was evaluated using the checklists of the National Institute for Health and Clinical Excellence (2009). The main criteria for the quality appraisal were (a) a clear statement of the aims of the research, (b) an appropriate research design, (c) a well-described and appropriate sampling strategy, data collection, and analysis method; and (d) a clear description of the research findings. Each study was rated as either high, medium, or low quality. Following this critical appraisal, five studies were excluded due to their low quality, leaving 42 studies for analysis. Appendix A shows the results of the critical appraisal process.

4.3.Analysis of the literature

The analysis of the literature was conducted under the guidelines proposed by Aveyard (2014). First, the main characteristics of the selected studies were summarized (see Appendix B). Then, every study was reread to identify and codify the relevant information. Consequently, all the studies were coded and assigned the following themes: operationalizations of SES (Theme 1); operationalizations of academic performance (Theme 2); relationship between SES and academic performance (Theme 3); prior academic achievement (Theme 4), university experience (Theme 5), and working status (Theme 6). The first research question was answered by summative content analysis. The second research questions were answered using both summative content analysis and meta-analysis.

4.3.1. Summative content analysis

A summative content analysis goes beyond quantifying the occurrence of words and content within texts; it aims at analyzing and interpreting that content (Hsieh & Shannon, 2005). To address the first research question, a summative content analysis of the first two themes listed above was carried out. Therefore, the operationalizations of SES (Theme 1) were categorized as parental educational level, parental occupation, income, household resources, or neighborhood resources. Similarly, the operationalizations of academic performance (Theme 2) were categorized as achievement, competencies, or persistence.

To answer the first part of the second research question, the relationship between SES and academic performance in higher education (Theme 3) was classified into three categories: significant and positive; significant and negative; and not significant. Those categories were chosen as they are the three possible basic ways in which two variables are correlated.

Subsequently, to answer the second part of the second research question, the mediating roles of prior academic achievement (Theme 4), university experience (Theme 5), and working status (Theme 6) were explored following the rationale that will be explained subsequently. The interest was to determine the effect size and the significance level of the relationships indicated with numbers in Figure 1. That is, the relationship between SES and academic performance (1), once the mediator was also introduced into the same explanatory model; and, next, the relationship between the mediator and academic performance (2). As such, a larger and significant relationship between the mediator and academic performance would suggest mediation. It is worth noting, however, that not all the analyzed references provided information on the relationship between the SES and the mediator (3). Consequently, this third relationship could not be included in the analysis, so it was not possible to distinguish between complete and partial mediation.

4.3.2. Statistical meta-analysis

Hunter and Schmidt (1990) define a meta-analysis as the quantitative summary and analysis of different effect sizes retrieved from several various studies. To further extend the answer to the second research question, a meta-analysis was carried out to calculate the average effect size of (a) the relationship between SES and academic performance and (b) the relationship between the investigated mediators and academic performance.

Metric from expressing effect sizes. The effect size (ES) selected for this study was Pearson's correlation coefficient (r). Most of the included studies reported a standardized regression coefficient, which was converted into an r-value using the guidelines proposed by Peterson and Brown (2005). Although alternative methods for transforming standardized regression coefficients into either partial or semi-partial correlations exist (Aloe & Becker, 2012; Aloe & Thompson, 2013; Fernández-Castilla et al., 2019), there was not enough information in the primary studies to

carry out such conversions. In four studies (Birch & Miller, 2006; Delaney, Harmon, & Redmond, 2011; Guimarães & Sampaio, 2013; Win & Miller, 2005) the unstandardized regression coefficient was first converted into a standardized coefficient using the standard deviation of both the predicted and the predictor variable as suggested by Bowman (2012). In six studies (Arulampalam, Naylor, & Smith, 2004; Hansen & Mastekaasa, 2006; Smith, 2016; Walpole, 2003; Yanbarisova, 2015; Yao, Zhimin, & Peng, 2015) the reported odds ratio were converted into an r-value following the procedure proposed by Borenstein, Hedges, Higgins, and Rothstein (2009). In one study (Waqas, Abbasi, & Idrees, 2013) the reported t-value was first converted into a standard mean difference (d-value) and, then, this d-value was converted into an r-value, also according to the procedure suggested by Borenstein et al. (2009).

In total, 13 studies were left out of the meta-analysis as will be explained next. The standardized regression coefficient in four articles (Loehr, Almarode, Tai, & Sadler, 2012; Morlaix & Suchaut, 2014; Pedrosa et al., 2007; Rodríguez, Ariza, & Ramos, 2014) could not be calculated due to insufficient information. It is essential to notice that the inclusion of unstandardized regression coefficients could dramatically increase the average effect size. Similarly, in two of the analyzed studies (Frischenschlager, Haidinger, & Mitteraurer, 2005; Stratton & Wetzel, 2011), it was not possible to convert the reported odds ratio due to a lack of information. In five of the articles (Anderton et al., 2016; Gerken and Volkwein, 2000; Ifenthaler & Widanapathirana, 2014; Nguyen, 2016; Zheng, Saunders, Shelley, Mack, & Whalen, 2002) the effect size was not reported due to it not being significant. Finally, in two of the articles (Bahamón & Reyes Ruiz, 2014; Triventi, 2014), the relationship between SES and academic performance was described but not reported through a specific effect size.

Statistical independence. One crucial methodological aspect when performing a meta-analysis is to ensure that the effect sizes are independent among them. To achieve such independence, the average effect size was calculated in studies reporting more than one effect size. In this way, the sample on which it was based contributed only with one effect size to the analysis (Sirin, 2005). One article (De Clercq et al., 2013) reported two different effect sizes coming from two different data sets. Both effect sizes were included in the meta-analysis.

Combining effect sizes across studies. Another major important consideration when conducting a meta-analysis is to decide how to transform the different effect sizes retrieved from the examined studies so that a meaningful and valid aggregation can be made. In this meta-analysis, the effect sizes were converted using Fisher's transformation as recommended by Hedges and Olkin (1985). Fisher's transformation is a variance stabilizing transformation so that the r-value becomes independent from the population p-value (Aloe & Becker, 2012; Bowman, 2012).

Homogeneity analysis. The homogeneity among the effect sizes was analyzed using Hedge's Q test for homogeneity. This test is based on chi-square statistics with k-1 degrees of freedom, where k is the number of investigated effect sizes. A significant result suggests that effect sizes across the studies are heterogeneous, so further exploration of the existence of possible mediators should be conducted.

Publication bias. A common fact in scientific literature is publication bias. Publication bias means that only statistically significant results or results which support the expected relationship are published (Rothstein, Sutton, & Borenstein, 2005). In this study, publication bias was assessed using the funnel plot. To evaluate the funnel plot symmetry, Egger's regression test (Egger, Smith, Schneider, & Minder, 1997) was performed, and the results indicated that there was no publication bias in the studies summarized through this meta-analysis ($z=.79$, $p = .43$).

Analysis of the mediators. Similar to the summative content analysis, the existence of mediators was analyzed by contrasting the effect size of the relationships indicated with numbers in Figure 1. The average effect size between SES and academic performance (1) was compared to the average effect size between the presumed mediator and academic performance (2). As such, a larger and significant average effect size between the mediator and academic performance would suggest mediation.

5. Results

The results of this systematic literature review are presented in accordance with the research questions. Sections 5.1 and 5.2 are based on the results derived from the summative content analysis, while Section 5.3 is based on the results of both the summative content analysis and the meta-analysis.

5.1.Measuring SES

Table 2 presents the classification of the operationalizations of SES into five major measurements: parental educational level, parental occupation, income, household resources, and neighborhood resources. Table 2 also provides information regarding the type of measure (i.e., single or composite score) and the type of scale (i.e., categorical or continuous) of each one of the examined operationalizations. It is important to note that a study can appear more than once in Table 2, as it might have reported several indicators for the assessment of SES. The classification of the operationalizations of SES is described in detail in the following sections.

5.1.1. Parental educational level

This category (n = 25) comprises operationalizations that assessed parental educational level as a single indicator (n = 21) or within a composite score (n = 4). With regard to "*single*

indicator" (*n* = 21), the use of categorical scales (*n* = 19) was identified to assess parental educational level in three possible ways: (a) the highest level of parental education (*n* = 4); (b) a dichotomous variable which indicated whether the parents had attended college (*n* = 3); and using several categories—ranging from no level of education to a university degree —to measure (c) both parents' level of education (*n* = 9); (d) only the mother's level (*n* = 2); and (e) only the father's level. Additionally, continuous scales (*n* = 2) were utilized to measure the length of the parents' education in two different ways. Firstly, Rothstein (2004) used the average number of years of education of the students' parents. Secondly, Delaney et al. (2011) converted the qualifications reported by the parents into years of education by estimating the number of years which are required to obtain those degrees. Consequently, this variable ranges from 8 (time necessary to complete primary school) to 19 (time necessary to complete a Ph.D.).

Regarding "*within a composite score*" (*n* = 4), parental educational level was assessed through both categorical (*n* = 3) and continuous scales. On the categorical scales (*n* = 3), De Clercq et al. (2017), and Rodríguez et al. (2014), reported the use of the highest level of parental education (*n* = 2). In addition, Gouvias, Katsis, and Limakopoulou (2012) measured parental education using the International Standard Classification of Education (ISCED). The ISCED is a classification of educational programs based on levels and areas of knowledge proposed in 1997 by the United Nations Educational, Scientific, and Cultural Organization (UNESCO), and then revised in 2011. The main goal behind this classification is the need to compare internationally different educational systems (Schneider, 2013). More specifically, there are nine ISCED levels ranging from early childhood education to a doctoral or equivalent level. On the continuous scales, Walpole (2003) recoded the parental educational level into a continuous SES variable based on the occupational prestige and socioeconomic scores as proposed by Nakao and Treas (1994). In

particular, Nakao and Treas (1994) suggested the following regression equation to calculate a socio-economic index (SEI): SEI = 9.24 + 0.64 (Education) + 0.31 (Income).

5.1.2. Parental occupation

This category (n = 10) includes operationalizations of parental occupation as a single indicator (n = 7) or within a composite score (n = 3). Regarding "*single indicator*" (n = 7), the use of categorical scales (n = 7) was identified to assess parental occupation in three possible ways: (a) as a dichotomous variable to indicate whether the parents worked; (b) as a dichotomous variable to indicate whether the parents' occupation(s) required a higher educational degree; (c) as several scales ranging from unskilled worker/unemployed to managerial/professional (n = 4). In addition, Smith (2016) reported the use of the Standard Occupational Classification scale 2000 (SOC 2000). The SOC 2000 is a classification system of the occupational structure of the United States, proposed by The Bureau of Labor Statistics. The SOC 2000 allows for comparisons to be made among paid occupations, based on the type of work, required skills, education, and training. More specifically, the SOC 2000 includes occupations from the public, private, and military sectors (Bureau of Labor Statistics, 2006).

Regarding "*within a composite score*" (n = 3), Rodríguez et al. (2014) used nominal categories which described parental occupation. In addition, Gouvias et al. (2012) reported the use of seven categories which ranged from unemployed to professionals, managers, and business owners. Finally, Walpole (2003) utilized the occupational prestige and socioeconomic scores, proposed by Nakao and Treas (1994), to recode parental occupation into a continuous SES variable.

5.1.3. Income

This category (*n* = 20) sorts between operationalizations that measured income as a single indicator (*n* = 17) or within a composite score (*n* = 3). Regarding "*single indicator*" (*n* = 17), the use of categorical scales (*n* = 16) was identified to measure income with: (a) several categories to discriminate between low and high income (*n* = 9); (b) dichotomous scales to determine whether students received financial aid (*n* = 4) or received subsidized lunches (*n* = 2); and (c) quartiles to classify university tuition fees. Additionally, Hansen and Mastekaasa (2006) proposed a continuous scale between 0 and 1 to determine the relative position of a student's family within the distribution of all family incomes (i.e., a score of 0.52 indicates that 52% of the families have less income).

Concerning "*within a composite score*" (*n* = 3), continuous scales were used by Rodríguez et al. (2014) to combine tuition fees and monthly income into a composite score. Similar to parental educational level and occupation, parental income was also re-coded by Walpole (2003) into a continuous SES variable, based on the occupational prestige and socioeconomic scores proposed by Nakao and Treas (1994).

5.1.4. Household resources

This category (*n* = 8) contains operationalizations of household resources as a single indicator (*n* = 6) or within a composite score (*n* = 2). Regarding "*single indicator*" (*n* = 6), categorical scales (*n* = 4) were selected to determine whether students have a computer and internet connection (*n* = 3); and books related to schoolwork. In addition, continuous scales (*n* = 2) were used to calculate household crowding. Harb and El-Shaarawi (2007) defined household crowding as the proportion of the number of family members to the number of rooms in the house.

Concerning "*within a composite score*" (*n* = 2), continuous scales were used to include the number of possessions as an indicator of household resources.

5.1.5. Neighborhood resources

This category (*n* = 9) presents operationalizations of socio-economic conditions at the area level. The use of both a single indicator (*n* = 3) and a composite score (*n* = 6) was identified in the examined primary studies. Among "*single indicator*" (*n* = 3), categorical scales (*n* = 3) were used to assess the students' stratum (*n* = 2) and degree of urbanization. In particular, the students' stratum refers to the six categories used by the Colombian government to classify households based on their physical characteristics and surroundings. The main reason behind this classification is to establish the price of public services hierarchically in each area (The World Bank, 2012).

Regarding "*composite score*" (*n* = 6), the connection of students' postal codes to several indices aimed at classifying areas based on socio-economic advantages and disadvantages was reported in the analyzed primary studies, with these indices being the Socio-Economic Indices for Areas (SEIFA) first developed by the Australian Bureau of Statistics. In particular, the Index of Economic Resources (Birch & Miller, 2006; Win & Miller, 2005); the Index of Education and Occupation (Win & Miller, 2005); and the Index of Relative Socioeconomic Advantage and Disadvantage (Puddey & Mercer, 2014) were identified in the examined references. Additionally, Thiele, Singleton, Pope, and Stanistreet (2016) used indices suggested by the Higher Education Funding Council for England (HEFCE). More specifically, the Index of Multiple Deprivation (IMD) was utilized to rank area deprivation in five quintiles (where the first quintile included the most deprived areas, and the fifth quintile included the least deprived areas). Finally, Shulruf, Hattie, and Tumen (2008) reported the use of the school decile, which is a system from New Zealand aimed at classifying schools into ten categories according to their students' socio-

economic conditions. Such a classification makes it possible to determine, for instance, whether schools are receiving students coming from low-SES areas.

5.1.6. Model for measuring SES

Figure 2 shows the model to assess SES resulting from the summative content analysis carried out in section 5.1. Education, occupation, income, household resources, and neighborhood resources are the five major measurements for assessing SES. Moreover, such measurements can be assessed at different possible levels, namely, the individual, family, or area levels.

5.2.Measuring academic performance in higher education

Table 3a and Table 3b present the classification of the operationalizations to assess academic performance in higher education into three measurements: achievement, competencies, and persistence, both before and during university. Table 3a and Table 3b also display information regarding the level of measurement (i.e., categorical or continuous) as well as the scale used in each of the examined references. Again, the studies can appear more than once in Table 3a and Table 3b, as they might have used several indicators to assess academic performance. The classification of the operationalizations of academic performance in higher education is explained in the upcoming sections.

5.2.1. Academic performance at university

This category includes operationalizations which assessed academic performance at university (n = 55). Operationalizations conforming to this category were grouped into three subcategories: academic achievement (n = 43); competencies (n = 6), and persistence (n = 6). Regarding the subcategory *academic achievement* (n = 43), the operationalizations identified in the analyzed studies were Grade Point Average (GPA; n = 32); single grades (n = 4); number of

credits (n = 6) and the relative achievement index. More specifically, three different types of GPA were identified in the examined references: the average at the end of the first year (n = 16), the current average when the study was conducted (n = 8), and the final average upon graduation (n = 8). Although both categorical (n = 8) and continuous (n = 24) scales were used to assess GPA, the most frequently used scales for measuring GPA were continuous from 0 to 4 (n = 15) and from 0 to 100 (n = 5). In addition, the use of a single grade was identified as the grades in subject areas (n = 3) and the grades on a first-year test. In the case of the number of credits (n = 6), Rodríguez and Ruiz (2011) proposed an original indicator —the degree progression index—which is the relationship between the actual number of credits, and the expected number of credits that a student must take. Moreover, Triventi (2014) suggested that ECTS credits are a concise way to measure students' academic achievement because they represent not only how many exams a student has completed satisfactorily but also their importance. Finally, the relative achievement index was proposed by Pedrosa et al. (2007) to compare students' performance at the entrance and exit of all courses. Differences in both the grading system and the number of students enrolled in the courses were the main reasons for developing this index, instead of using the actual numerical value of the grades.

In the subcategory *competencies* (n = 6), Bahamón and Reyes Ruiz (2014) and Rodríguez et al. (2014) used the SABER PRO test, which is a standardized test taken by the students at the end of the university level in Colombia to determine both their general competencies (e.g., quantitative reasoning, critical reading) and specific competencies within each field of study. Similarly, students' proficiency in English was assessed through several standardized tests such as the College English Test (CET) and the internal Post-Entrance Literacy Assessment (PELA). In addition, Morlaix and Suchaut (2014) reported the use of the *Diplôme Approfondi de Langue*

Française (DALF) test to evaluate the first-year students' written comprehension of French. Finally, Puddey and Mercer (2014) reported the use of the Graduate Medical School Assessment Test (GAMSAT) to assess the level of preparation for undertaking studies in medicine and/or attending medical school at the graduate level at Australian, British, and Irish universities.

Regarding the subcategory *persistence* (*n* = 6), all the analyzed operationalizations were related to completion status. As such, the operationalizations grouped in this subcategory were dropout (*n* = 2), final degree classification (*n* = 2), attending graduate school, and graduation rate. In particular, the UK undergraduate degree classification was sorted into this subcategory.

5.2.2. Academic performance before university

This category includes operationalizations which assessed academic performance before attending university (*n* = 43). In this case, the analyzed operationalizations were also grouped in three subcategories: prior academic achievement (*n* = 20), competencies (*n* = 15), and persistence (*n* = 8).

In the subcategory *prior academic achievement* (*n* = 20), academic performance before university was measured using HSGPA (*n* = 14), single grades (*n* = 3), and subject area exams (*n* = 3). HSGPA was calculated through both categorical (*n* = 4) and continuous (*n* = 10) scales. In particular, the continuous scale ranging from 0 to 4 (*n* = 4), was the most frequently reported in the analyzed studies. Regarding single grades, grades in specific subjects such as science, mathematics, and English (*n* = 2) alongside biology, chemistry, and physics, were reported in the examined primary studies. Furthermore, subject area exams such as the NCEA qualification system in New Zealand (Shulruf et al., 2008), the Leaving Certificate in Ireland (Delaney et al.,

2011), and the SABER 11 test in Colombia (Bahamón & Reyes Ruiz, 2014) were identified in the examined studies.

In the subcategory *competencies* (*n* = 15), several entrance exams to university were identified in the analyzed studies. These exams were the SAT (formerly known as the Scholastic Assessment Test; *n* = 9) and the ACT (originally American College Testing; *n* = 3) in the USA. These entrance exams are of importance because they indicate students' academic preparation for university. Remarkably, in three of the investigated references (Gouvias et al., 2012; Guimarães & Sampaio, 2013; Nguyen, 2016), *subject area exams* were used as entrance exams to the university.

Within the subcategory *persistence* (*n* = 8), the operationalizations of academic performance before university were grouped between admission ranks (*n* = 5) and grade retention (*n* = 3). In relation to admission ranks, the Tertiary Entrance Rank (TER) and the Australian Tertiary Admission Rank (ATAR) were identified in the analyzed studies. ATAR ranks students' previous academic achievement in high school and is mostly used as an admission criterion for university in Australia (Li & Dockery, 2014). It is important to note that the ATAR replaced the TER in 2010. Regarding grade retention (*n* = 3), the operationalizations identified in the analyzed studies were the years repeated in high school (*n* = 2) and the number of class repetitions.

5.2.3. Model for measuring academic performance in higher education

Figure 3 proposes a model to measure academic performance in higher education drawn from the summative content analysis conducted in section 5.2. Achievement, competencies, and persistence are the three measurements used to assess academic performance in higher education. Furthermore, those measurements can be assessed both during and before university.

5.3.The relationship between SES and academic performance in higher education

Table 4 provides information on the relationship between SES and academic performance in higher education, as well as the mediators identified in the analyzed primary studies. Such information was analyzed through both summative content analysis and meta-analysis. This section first presents the results for the SES-academic performance relationship. Next, the results for the investigated mediators are provided.

The summative content analysis demonstrated that there are three types of relationships between SES and academic performance in higher education: namely, positive (n = 25), negative (n = 6), and no significant (n = 12) relationships. A *positive relationship* indicated that the better one's socio-economic conditions, the better one's academic performance in higher education. However, a closer review of the *negative relationship* (n = 6) revealed enlightening information. Pedrosa et al. (2007) indicated that students who came from public schools had a better academic performance than their counterparts coming from private schools. Those students with less favorable socio-economic conditions were able to develop a certain "educational resilience", which was described by Pedrosa et al. (2007) as the process of transforming early disadvantages in life into better academic performance in higher education. In addition, students who either (a) were scholarship holders (Morlaix & Suchaut, 2014), (b) lived in a crowded house (Harb & El-Shaarawi, 2007), or (c) came from high schools with a high proportion of free/reduced lunches (Black, Lincove, Cullinane, & Veron, 2015), also had low academic performance in university. Finally, Arulampalam et al. (2004) suggested that students who had higher scores in biology, chemistry, and physics in high school also had a lower likelihood of dropping out during university.

The meta-analysis revealed that the average effect size of the relationship between SES and academic performance in higher education was weak and significant (ES = .06, Se = .013, CI = [.03; .08], $p < .001$). The Q test for homogeneity for this average effect size was significant (Q = 460.30, df = 22, $p < .001$). Figure 4 displays the forest plot with the average effect size of the relationship between SES and academic performance in higher education.

5.3.1. Prior academic achievement

The mediating role of prior academic achievement was investigated based on the information presented in Table 4 and following the rationale explained in section 4.3.1. Therefore, it was possible to compare the relationship between SES and academic performance with the relationship between prior academic achievement and academic performance.

The summative content analysis suggested that HSGPA ($n = 15$), entrance exams ($n = 5$), admission ranks ($n = 3$), failure in secondary school ($n = 2$), and type of high school diploma were the potential mediators of the relationship between SES and academic performance in higher education. Regarding *HSGPA* ($n = 15$), high school grades were more strongly related to academic performance than SES. In such cases, the relationship between SES and academic performance was (a) positive ($n = 9$), (b) negative ($n = 2$), and (c) not significant ($n = 4$). Concerning *entrance exams* ($n = 5$), it was found that SAT scores ($n = 4$) and Leaving Certificate scores were more strongly related to academic performance than SES. Correspondingly, the reported relationship between SES and academic performance was (a) positive ($n = 4$) and (b) not significant. In addition, *failure in secondary school* ($n = 2$) was more strongly related to academic performance than SES. In such cases, the relationship between SES and academic performance was (a) negative and (b) not significant. Finally, Harb and El-Shaarawi (2007) indicated that receiving a scientific diploma in high school was more strongly related to GPA than living in a crowded household.

The meta-analysis demonstrated that the average effect size of the relationship between prior academic achievement and academic performance in higher education was positive and significant (ES = .29, Se = .07, CI = [.15; .42], $p < .001$). The Q test for homogeneity for this average effect size was significant (Q = 12088.23, df = 20, $p < .001$).

5.3.2. University experience

The mediating role of university experience was investigated based on the information presented in Table 4 and following the rationale explained in section 4.3.1. Thus, the relationship between SES and academic performance was compared to the relationship between university experience and academic performance.

The summative content analysis revealed that academic experience ($n = 3$) and institutional experience ($n = 2$) were the potential mediators of the relationship between SES and academic performance in higher education. Regarding *academic experience* ($n = 3$), Bruinsma and Jansen (2007) indicated that students who had higher grades were also more satisfied with the teacher's ability to explain the topic, the teacher's ability to use the resources, and the teacher's openness to questions. Moreover, the higher the study load, the lower the grades in the course would be, while the higher the number of contact hours, the higher the grades. In this case, Bruinsma and Jansen (2007) also found no significant relationship between SES and academic performance in higher education. Similarly, Walpole (2003) suggested that low-SES students who had worked on their professors' research during university were more likely to enroll in graduate programs following their completion of their undergraduate degree at university. In this case, the relationship between SES and academic performance was significant but weaker. Finally, Gerken and Volkwein (2000) identified that students' academic conscientiousness was more strongly related to their degree completion and GPA than their parental educational level. Regarding *institutional experience* ($n =$

2), Harb and El-Shaarawi (2007) found that students who had a positive attitude towards university also performed better academically. This influence was greater than the influence of SES on academic performance. Also, Wang, Kong, Shan, and Vong (2010) reported that students' sense of belonging was more strongly related to students' GPA than their father's education and family income.

The meta-analysis revealed that the average effect size of the relationship between university experience and academic performance in higher education was positive and significant (ES = .13, Se = .02, CI = [.09; .16], $p < .001$). The Q test for homogeneity for this average effect size was not significant (Q = 2.4094, df = 5, p = .79).

5.3.3. Working status

The mediating role of working status was investigated based on the information presented in Table 4 and following the rationale explained in section 4.3.1. Hence, a comparison was made between the relationship between SES and academic performance and the relationship between working status and academic performance.

The summative content analysis indicated that whether a student was employed or not (*n* = 5), the characteristics of their work (*n* = 2) and their reasons for working were the potential mediators of the relationship between SES and academic performance in higher education. Regarding *being employed or not* (*n* = 5), the influence of employment on academic performance was stronger than the influence of SES. In such cases, the relationship between SES and academic performance was (a) positive (*n* = 3), (b) negative, and (c) not significant. With respect to the *characteristics of work*, Yanbarisova (2015) found that students who were working full-time outside their academic fields displayed worse academic performance than their counterparts. In

this case, the relationship between SES and academic performance was not significant. In addition, Wang et al. (2010) indicated that jobs that provided students the opportunity to learn new things had a greater influence on their academic performance than their father's education and occupation, and their family income. Regarding the *reasons for working*, Wang et al. (2010) also indicated that when the reason for working was to acquire working experience, the part-time jobs were more strongly related to students' GPA than their father's education and occupation, and their family income.

The meta-analysis demonstrated that the average effect size of the relationship between working status and academic performance in higher education was negative and significant (ES = −.10, Se = .05, CI = [-.19; -.01], $p < .001$). The Q test for homogeneity for this average effect size was significant (Q = 51.7643, df = 7, $p < .001$).

In summary, the summative content analysis suggested that the investigated mediators were more strongly related to academic performance than SES. Furthermore, the meta-analysis showed that the average effect sizes of the mediators were significant and larger than the effect size of the SES-academic performance relationship. The mediator with the largest average effect size was prior academic achievement, followed by university experience, and working status.

6. Discussion

The objectives of this systematic literature review were (1) to analyze how SES and academic performance in higher education are measured; (2) to determine whether the relationship between SES and academic performance in higher education is mediated by a) prior academic achievement; b) university experience; and c) working status.

6.1. Conclusions and implications for practice

6.1.1. Measuring SES

The first conclusion of this study is that five major measurements should be considered when assessing SES: education, occupation, income, household resources, and neighborhood resources. The findings of this systematic literature review also suggest specific ways to operationalize each one of the measurements of SES.

First, it was found that education is traditionally assessed through categorical variables that indicate the achieved academic degree, ranging from no education up to a doctoral degree. In this respect, the use of the International Standard Classification of Education (ISCED) is highly recommended. The ISCED has established a unique scale which allows for comparisons to be made among different international contexts.

Second, the results of this study showed that occupation is predominantly assessed using categorical scales. In this regard, a well-established classification to measure occupation is the Standard Occupation Classification (SOC). Such a classification is based on the type of work, skills, and level of education. Even though the SOC was first proposed in the USA, national variants also exist in countries within Europe and Asia. Therefore, the SOC could be used not only for international comparisons but also as a classification system for countries which have not developed their own classification system.

Third, the findings of this study indicated that income is typically measured with intervals to categorize the amount of money being earned. However, a more advisable way to operationalize income is through a multiple of the minimum wage paid in each country (i.e., equivalent to one minimum wage, twice minimum wage, etc.). As such, the minimum wage could be related to the

type of work and level of education of a wage-earner. Besides, the use of the minimum wage as a measurement would allow for the comparison of socio-economic conditions across several different countries.

Fourth, although the results from this study revealed that measures of household resources are related to the possessions available at home, it seems adequate to distinguish between material resources and cultural resources (Gouvias et al., 2012). Material resources are merely the items students have at home (i.e., a personal computer, internet connection, an individual room), while cultural resources are items which might represent an intellectual added value to the students (i.e., books related to schoolwork or musical instruments).

Finally, this study revealed that neighborhood resources can be operationalized through indexes which rank areas according to their socio-economic advantages and disadvantages. Well-established examples of these indexes are the Socio-Economic Indexes for Areas (SEIFA) proposed by the Australian Bureau of Statistics and the Index of Multiple Deprivation (IMD) proposed by the UK government. Even though these indexes were developed for specific national contexts, the methodology underlying their creation can be replicated to develop indexes within each country. However, it does not seem convenient to use these indices directly to measure students' SES. What is recommended, instead, is to use area-based indicators such as SEIFA to achieve a better understanding of the social and economic conditions where students live (Australian Bureau of Statistics, 2011). This suggestion is consistent with previous meta-analytic studies focusing on primary and secondary education (e.g., Sirin, 2005; White, 1982), where it has been recommended to avoid using aggregated indicators to assess SES at the individual level, as they can overestimate the effect of the relationship between SES and academic performance.

6.1.2. Measuring academic performance in higher education

The second conclusion of this study is that academic performance in higher education should be assessed considering three major measurements: academic achievement, competencies, and persistence. This study also identified several ways to operationalize such measures both during and prior to university.

Regarding academic achievement, the results from this study corroborate the assertion that first-year GPA is the most used operationalization of academic achievement at university, as first-year GPA is considered a strong predictor of subsequent academic outcomes at university (Cliffordson, 2008; Gerken & Volkwein, 2000). When the interest is to compare GPA across different fields and institutions, it is necessary to take into account that the grading system does change between fields and institutions of higher education; furthermore, there could also be variations in the assessment process regardless of the use of the same grading scale (Hansen & Mastekaasa, 2006). In addition, the findings of this study also reveal that HSGPA is the most common operationalization of prior academic achievement before university. HGSPA depends on the curriculum followed in each institution (Westrick et al., 2015), the quality and strictness of the scoring system, as well as the student population in each institution. The selection of HSGPA as a predictor of academic performance increases the explained variance of GPA at university (Zheng et al., 2002), and furthermore, HSGPA seems to have a larger predictive validity than entrance exams, regardless of the grading system and the academic program (Cliffordson, 2008).

With reference to competencies, the findings of this systematic literature review indicate that competencies have been operationalized through the outcomes of standardized testing both during and prior to university. However, the difference across levels might lie on the purpose of such tests. In the case of academic performance at university, standardized tests aim to evaluate

the acquisition of both general and specific competencies pertaining to each study area. In the case of academic performance before enrolling in university, standardized tests are designed as entrance exams to estimate the students' academic preparation for their university studies.

Nevertheless, the use of achievement tests as entrance exams to university was also identified in the examined primary studies. This finding is somewhat problematic for several reasons. Firstly, achievement tests are designed to measure past accomplishment in learning instead of measuring the capacity for future accomplishment (Sidhu, 2005). Secondly, whereas a great deal is known about the predictive validity of aptitude tests (e.g., the SAT I), the predictive validity of achievement tests (e.g., the SAT II) is still unclear in the literature (Cliffordson, 2008). Thirdly, Zwick (2012) has suggested that achievement tests might indicate the degree to which wealthier students have access to either better information for the test (content hypothesis) or better preparation for the test (coaching hypothesis). Therefore, achievement tests seem to be more related to students' SES than aptitude tests.

Finally, this study suggests that persistence in university can be measured in terms of students' degree completion. In addition, persistence before university can be assessed as the students' academic rank and grade repetition. An important conceptual distinction between persistence and retention should be acknowledged, as persistence is an individual phenomenon, while retention is an institutional one; therefore, these terms should not be used interchangeably (Reason, 2009).

6.1.3. The relationship between SES and academic performance in higher education

The third conclusion of this study is that the relationship between SES and academic performance in higher education is weak. This result is coherent with previous meta-analyses in

higher education, which have also reported a weak effect size (e.g., Richardson et al., 2012). A critical interpretation of the findings of this study could beg the question of the actual importance of SES as a predictor of academic performance in higher education. To begin with, it could be the case that the influence of socio-economic conditions on students' performance is lower in higher education than in previous levels of education (the influence of SES on prior academic achievement has been well-established in educational research). One can also argue that regardless of the hindrances that low-SES students face when entering university, those who are admitted share a similar educational experience to their wealthier counterparts (Smith, 2016). Therefore, the higher education system could have the same influence on any student despite his or her socio-economic conditions.

Furthermore, the findings of this study indicate that the relationship between SES and academic performance in higher education is weak when other factors are considered. This fact urges the educational research field to select more robust analytical techniques when investigating academic performance in higher education. A mere bivariate analysis no longer suffices. In addition, regardless of the multiple theories in higher education which suggest that SES and academic performance are positively related, strong empirical evidence supporting these theoretical claims is still missing in the educational literature (Marks, 2017).

However, a weak relationship between SES and academic performance in higher education does not imply that low-SES students should be ignored or that increasing their participation in university should be dismissed. How to efficiently attend to low-SES students' educational needs remains a challenge for the higher education system, and in this respect, the findings of this systematic literature review might be transferred in one of three ways to properly deal with that challenge.

Firstly, it was identified that prior academic achievement is more strongly related to academic performance in university than SES. While this finding is not surprising, it does support the need to reinforce the past performance of low-SES students through, for instance, academic preparation courses before university. An example of this type of program is the enabling programs proposed by the Australian government. Enabling programs are designed to provide disadvantaged students with specific competencies (e.g., literacy, numeracy, communication, and critical thinking skills) so that they can be prepared for their university studies (Pitman et al., 2016). Moreover, enabling programs are effective pathways to higher education for almost half of the enrolled disadvantaged students (Hodges et al., 2013).

Secondly, the findings from this study also revealed that the influence of university experience on academic performance is larger than the influence of SES. The defining factors of the university experience such as classroom climate, quantity and quality of the instruction (Bruinsma & Jansen, 2007), sense of belonging and school integration (Wang et al., 2010), and peer support (De Clercq et al., 2013) would help low-SES students to adapt to their new academic settings at university. Interestingly, Devlin, Kift, Nelson, Smith, and McKay (2012) have proposed a set of teaching guidelines in order to foster low-SES students' academic performance. Far from being prescriptive, these guidelines can be understood as key practical advice for teachers whose students come from low socio-economic settings.

Thirdly, there was also evidence of the likely mediating role of working status in the SES-academic performance relationship. This finding supports the assumption that working during university studies might have a negative influence on students' performance. However, students who work part-time within their academic fields might perform better academically than students who work outside their academic areas (Wang et al., 2010; Yanbarisova, 2015). Thus, working

should allow low-SES students not only to overcome their financial needs but also to extend their academic experience by increasing their body of knowledge while attending university. A concrete example of this type of job for low-SES students is serving as an undergraduate teaching assistant (UTA).

6.2.Limitations

Although the results of this systematic literature review provide insights into the relationship between SES and academic performance in higher education, several limitations of the present study should be acknowledged. First, the number of studies which explore the relationship between SES and academic performance is quite low. Hence, all studies that have aimed to predict academic performance in higher education were considered. However, a precise definition of the relationship between SES and academic performance was often lacking in the analyzed studies. Second, several additional variables which might also interact with SES were not always reported in the reviewed studies. Thus, variables such as students' cognitive factors could not be considered for the summative content analysis. Third, information about the students' academic programs was not always reported in the analyzed studies. Therefore, whether the relationship between SES and academic performance in higher education depends on students' academic programs could not be determined. Fourth, there was no information on the relationship between SES and the meta-analyzed mediators. Therefore, a complete analysis of the selected mediators could not be carried out. Finally, the Q test for homogeneity was not significant for the average effect size of university experience. This result might be suggesting that the resulting average effect size is less generalizable.

6.3.Implications for future research

Starting from the findings and limitations of this systematic literature review discussed earlier, future research in higher education could benefit from focusing on several topics, listed as follows. Firstly, it is interesting to note that using dimensional reduction techniques such as Principal Component Analysis (PCA) leads to a composite score to assess SES. However, composite scores also represent a limitation to fully capturing the underlying variance of SES indicators. Therefore, the matter of how to construct composite scores to assess SES is an issue which requires further exploration.

Secondly, previous research at the elementary and secondary educational levels (e.g., Sirin, 2005; White, 1982) has demonstrated the moderating role of methodological aspects, such as the type of SES-achievement measure, on the relationship between SES and academic performance. However, exploring such a moderating role still remains an unfinished task for the higher education field.

Thirdly, the criticism that standardized tests merely measure students' socio-economic conditions and do not predict their future academic performance in higher education (Mattern, Shawn, & Williams, 2008) persists among some educational researchers. Therefore, further research could contribute to clarifying the relationship between standardized test outcomes, SES, and academic performance at university.

Fourthly, when considering working status to predict academic performance, it seems essential to include additional characteristics such as the type of work (part-time or full-time), correspondence with the academic field, and reasons for working. In fact, including working status merely as a dichotomous variable in any predictive analysis of academic performance could be

misleading (Wang et al., 2010). Therefore, to gain a better understanding of the influence of working status, both quantitative and qualitative research methods are highly recommended (Yanbarisova, 2015).

Fifthly, after completion of their undergraduate degree, low- SES students are more likely to join the workforce instead of pursuing a postgraduate degree, as seems to be the case with high-SES students (Walpole, 2003). However, an additional question worth exploring is what happens with the relationship between SES and the academic performance of low- SES students who do continue to the postgraduate educational level.

Finally, recent educational research (Musso, Kyndt, Cascallar, & Dochy, 2013, 2012; Cascallar, Musso, Kyndt, & Dochy, 2015; Kyndt, Musso, Cascallar, & Dochy, 2015; Musso & Cascallar, 2009) has utilized predictive systems based on artificial neural network approaches to study academic performance in primary, secondary, and higher education. The improvement of the validity, the increase in the accuracy of the predictions and classifications, and the possibility to determine the predictive weight contributions of each of the variables in the models are the principal advantages of developing predictive systems based on artificial neural networks. Therefore, we would like to encourage the use of artificial neural networks in order to gain a more exhaustive understanding of the relationship between SES and academic performance in higher education.

References

Studies marked with an asterisk (*) were included in the summative content analysis.

Studies marked with a cross (+) were included in the meta-analysis.

Aloe, A. M., & Becker, B. J. (2012). An effect size for regression predictors in meta-analysis. *Journal of Educational and Behavioral Statistics*, *37*, 278-297. doi: 10.3102/1076998610396901

Aloe, A. M., & Thompson, C. G. (2013). The synthesis of partial effect sizes. *Journal of the Society for Social Work and Research*, *4*, 390-405. doi: 10.5243/jsswr.2013.24

*Anderton, R., Evans. T., & Chivers, P. (2016). Predicting academic success of health science students for first year anatomy and physiology. *International Journal of Higher Education*, *5*, 250-260. doi:10.5430/ijhe.v5n1p250

*+Arulampalam, W., Naylor, R., & Smith, J. (2004). Factors affecting the probability of first year medical student dropout in the UK: A logistic analysis for the intake cohorts of 1980–92. *Medical Education*, *38*, 492-503. doi:10.1046/j.1365-2929.2004.01815.x

Astin, A. W. (1984). Student involvement: A developmental theory for higher education. *Journal of College Student Personnel*, *25*, 297-308.

Astin, A. W. (1993). *What matters in college? Four years revisited*. San Francisco, USA: Jossey-Bass.

Astin, A. W. (1999). Student involvement: A developmental theory for higher education. *Journal of College Student Development*, *40*, 518-529.

Australian Bureau of Statistics. (2011). *Measures of socioeconomic status*. (ABS Publication No. 1244.0.55.001). Retrieved from: http://www.ausstats.abs.gov.au/Ausstats/subscriber.nsf/0/367D3800605DB064CA2578B60013445C/$File/1244055001_2011.pdf

Aveyard, H. (2014). *Doing a literature review in health and social care: A practical guide*. London, United Kingdom: McGraw-Hill Education. doi:10.7748/nr.18.4.45.s2

*Bahamón, M. J., & Reyes Ruiz, L. (2014). Caracterización de la capacidad intelectual, factores sociodemográficos y académicos de estudiantes con alto y bajo desempeño en los exámenes SABER PRO-AÑO 2012. [Characterization of the intellectual capacity, sociodemographic, and academic factors of the students with high and low performance in the SABER PRO test – 2012]. *Avances en Psicología Latinoamericana, 32*, 459-476. doi:10.12804/apl32.03.2014.01

Betts, L. R., Elder, T. J., Hartley, J., & Blurton, A. (2008). Predicting university performance in Psychology: The role of previous performance and discipline-specific knowledge. *Educational Studies, 34*, 543-556. doi:10.1080/03055690802288528

*+Beyene, K. M., & Yimam, J. A. (2016). Multilevel analysis for identifying factors influencing academic achievement of students in higher education institution: The case of Wollo University. *Journal of Education and Practice, 7*, 17-23.

*+Birch, E. R., & Miller, P. W. (2006). Student outcomes at university in Australia: A quantile regression approach. *Australian Economic Papers, 45*, 1-17. doi:10.1111/j.1467-8454.2006.00274.x

Black, S. E., Lincove, J., Cullinane, J., & Veron, R. (2015). Can you leave high school behind? Economics of Education Review, 46*, 52-63. doi:10.3386/w19842

Blömeke, S., Zlatkin-Troitschanskaia, O., Kuhn, C., & Fege, J. (2013). Modeling and measuring competencies in higher education. In Blömeke S., Zlatkin-Troitschanskaia O., Kuhn C., Fege J. (Eds.), *Professional and Vet Learning: Vol. 1. Modeling and Measuring Competencies in Higher Education* (pp. 1-10). Dordrecht, The Netherlands: Sense Publishers. doi:10.1007/978-94-6091-867-4

*+Bonsaksen, T. (2016). Predictors of academic performance and education programme satisfaction in occupational therapy students. *British Journal of Occupational Therapy*, *79*, 361-367. doi:10.1177/0308022615627174

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. Chichester, UK: John Wiley.

Bourdieu, P. (1986). The forms of capital. In J. Richardson (Ed.) *Handbook of theory and research for the sociology of education* (pp.241-258). New York, NY: Greenwood.

Bowman, N. A. (2012). Effect sizes and statistical methods for meta-analysis in higher education. *Research in Higher Education, 53*, 375-382. doi: 10.1007/s11162-011-9232-5

*+Bozick, R. (2007). Making it through the first year of college: The role of students' economic resources, employment, and living arrangements. *Sociology of Education, 80*, 261-285. doi: 10.1177/003804070708000304

*+Bruinsma, M., & Jansen, E. P. W. A. (2007). Educational productivity in higher education: An examination of part of the Walberg educational productivity model. *School Effectiveness and School Improvement, 18*, 45-65. doi:10.1080/09243450600797711

Bureau of Labor Statistics. (2006). *Standard Occupational Classification-Revision for 2010.* Retrieved from https://www.bls.gov/soc/soc_may06.pdf

Byrne, M., & Flood, B. (2008). Examining the relationships among background variables and academic performance of first year accounting students at an Irish University. *Journal of Accounting Education, 26,* 202-212. doi:10.1016/j.jaccedu.2009.02.001

Cascallar, E., Musso, M. F., Kyndt, E., & Dochy, F. (2015). Modelling for understanding and for prediction/classification the power of neural networks in research. *Frontline Learning Research, 2,* 67-81, doi:10.14786/flr.v2i5.135

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,* 407-420. doi:10.1037/a0027180

Cliffordson, C. (2008). Differential prediction of study success across academic programs in the Swedish context: The validity of grades and tests as selection instruments for higher education. *Educational Assessment, 13,* 56-75. doi:10.1080/10627190801968240

Coleman, J., Campbell, E., Hobson, C., McPartland, F., Mood, A., & Weinfeld, F. (1966). *Equality of educational opportunity*. Washington DC: U.S. Government.

Cowan, C. D., Hauser, R., Kominski, R., Levin, H., Lucas, S., Morgan, S., Spencer, M.B., & Chapman, C. (2012). *Improving the measurement of socioeconomic status for the national assessment of educational progress: A theoretical foundation*. Retrieved from National Center for Education Statistics website https://nces.ed.gov/nationsreportcard/pdf/researchcenter/Socioeconomic_Factors.pdf

Crawford, C. (2014). *Socio-economic differences in university outcomes in the UK: Drop-out, degree completion and degree class* (Institute for Fiscal Studies Working Paper No. W14/31). Retrieved from Institute for Fiscal Studies website https://www.ifs.org.uk/uploads/publications/wps/WP201431.pdf

*+De Clercq, M., Galand, B., Dupont, S., & Frenay, M. (2013). Achievement among first-year university students: An integrated and contextualized approach. *European Journal of Psychology of Education*, *28*, 641-662. doi:10.1007/s10212-012-0133-6

*+De Clercq, M., Galand, B., & Frenay, M. (2017). Transition from high school to university: A person-centered approach to academic achievement. *European Journal of Psychology of Education*, *32*, 39-59. doi:10.1007/s10212-016-0298-5

*+Delaney, L., Harmon, C., & Redmond, C. (2011). Parental education, grade attainment and earnings expectations among university students. *Economics of Education Review*, *30*, 1136-1152. doi:10.1016/j.econedurev.2011.04.004

Devlin, M., Kift, S., Nelson, K., Smith, E., & McKay, J. (2012). *Effective teaching and support of students from low socioeconomic status backgrounds: Resources for Australian higher education*. Sydney, Australia: Department of Industry, Innovation, Science, Research, and Tertiary Education.

Dika, S. L., & Singh, K. (2002). Applications of social capital in educational literature: A critical synthesis. *Review of Educational Research*, *72*, 31-60. doi:10.3102/00346543072001031

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ*, *315*, 629-634. doi: 10.1136/bmj.315.7109.629

Erola, J., Jalonen, S., & Lehti, H. (2016). Parental education, class and income over early life course and children's achievement. *Research in Social Stratification and Mobility*, *44*, 33-43. doi:10.1016/j.rssm.2016.01.003

Fernández-Castilla, B., Aloe, A. M., Declercq, L., Jamshidi, L., Onghena, P., Beretvas, S. N., & Van den Noortgate, W. (2019). Concealed correlations meta-analysis: A new method for synthesizing standardized regression coefficients. *Behavior Research Methods*, *51*, 316-331. doi: 10.3758/s13428-018-1123-7

*Frischenschlager, O., Haidinger, G., & Mitterauer, L. (2005). Factors associated with academic success at Vienna Medical School: Prospective survey. *Croatian Medical Journal, 46,* 58-65.

Galobardes, B., Shaw, M., Lawlor, D. A., Lynch, J. W., & Smith, G. D. (2006). Indicators of socioeconomic position (part 1). *Journal of Epidemiology & Community Health, 60,* 7-12. doi:10.1136/jech.2004.023531

Gamoran, A., & Long, D. A. (2007). Equality of educational opportunity. A 40-year retrospective. In R. Teese, S. Lamb, & M. Duru-Bellat (Eds.), *International studies in educational inequality, theory and policy* (pp. 23-47). Dordrecht, The Netherlands: Springer. doi:10.1007/978-1-4020-5916-2_2

Ganzeboom, H. B., De Graaf, P. M., & Treiman, D. J. (1992). A standard international socio-economic index of occupational status. *Social Science Research, 21,* 1-56. doi:10.1016/0049-089x(92)90017-b

*Gerken, J. T., & Volkwein, J. F. (2000, May). *Pre-college characteristics and freshman year experiences as predictors of 8-year college outcomes*. Paper presented at the annual meeting of the Association for Institutional Research, Cincinnati, OH.

Glasman, N. S., & Biniaminov, I. (1981). Input-output analyses of schools. *Review of Educational Research, 51,* 509-539. doi: 10.3102/00346543051004509

*+Gouvias, D., Katsis, A., & Limakopoulou, A. (2012). School achievement and family background in Greece: A new exploration of an omnipresent relationship. *International Studies in Sociology of Education, 22,* 125-145. doi:10.1080/09620214.2012.700186

*+Guimarães, J., & Sampaio, B. (2013). Family background and students' achievement on a university entrance exam in Brazil. *Education Economics*, *21*, 38-59. doi:10.1080/09645292.2010.545528

*+Hansen, M. N., & Mastekaasa, A. (2006). Social origins and academic performance at university. *European Sociological Review*, *22*, 277-291. doi:10.1093/esr/jci057

*+Harb, N., & El-Shaarawi, A. (2007). Factors affecting business students' performance: The case of students in United Arab Emirates. *Journal of Education for Business*, *82*, 282-290. doi:10.3200/joeb.82.5.282-290

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. London: Academic Press.

Heyvaert, M., Maes, B., & Onghena, P. (2013). Mixed methods research synthesis: definition, framework, and potential. *Quality & Quantity*, *47*, 659-676. doi: 10.1007/s11135-011-9538-6

Hilton, T. (1982). *Persistence in higher education: A empirical study*. (College Board report 82-5). Retrieved from College Board website http://research.collegeboard.org/sites/default/files/publications/2012/7/researchreport-1982-5-persistence-higher-education.pdf

Hodges, B., Bedford, T., Hartley, J., Klinger, C., Murray, N., O'Rourke, J., & Schofield, N. (2013). *Enabling retention: processes and strategies for improving student retention in university-based enabling programs: Final report 2013*. Sydney, Australia: Australian Government Office for Learning and Teaching.

Hsieh, H. F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research*, *15*, 1277-1288. doi:10.1177/1049732305276687

Hunter, J. E., & Schmidt, F. L. (1990). *Methods of meta-analysis. Correcting error and bias in research findings*. California: Sage Publications.

*Ifenthaler, D., & Widanapathirana, C. (2014). Development and validation of a learning analytics framework: Two case studies using support vector machines. *Technology, Knowledge and Learning*, *19*, 221-240. doi:10.1007/s10758-014-9226-4

Jæger, M. M. (2011). Does cultural capital really affect academic achievement? New evidence from combined sibling and panel data. *Sociology of Education, 84,* 281-298. doi:10.1177/0038040711417010

Kelly, L. J. (1996, May). *Implementing Astin's IEO model in the study of student retention: A multivariate time dependent approach*. Paper presented at the 36th Annual Forum of the Association for Institutional Research, Albuquerque, NM.

Krieger, N., Williams, D. R., & Moss, N. E. (1997). Measuring social class in US public health research: Concepts, methodologies, and guidelines. *Annual Review of Public Health*, *18*, 341-378. doi:10.1146/annurev.publhealth.18.1.341

Kyndt, E., Musso, M., Cascallar, E.C., & Dochy, F. (2015). Predicting academic performance: The role of cognition, motivation and learning approaches. A neural network analysis. In V. Donche, S. De Maeyer, D. Gijbels, & H. van den Bergh (Eds.), *Methodological challenges in research on student learning* (pp. 55-76). Antwerp, Belgium: Garant.

Li, G., Chen, W., & Duanmu, J. L. (2010). Determinants of international students' academic performance: A comparison between Chinese and other international students. *Journal of Studies in International Education, 14*, 389-405. doi:10.1177/1028315309331490

Li, I. W. & Dockery, A. M. (2014). *Socio-economic Status of Schools and University Academic Performance: Implications for Australia's Higher Education Expansion*. National Centre for Student Equity in Higher Education (NCSEHE), Perth: Curtin University.

*Loehr, J. F., Almarode, J. T., Tai, R. H., & Sadler, P. M. (2012). High school and college biology: A multi-level model of the effects of high school courses on introductory course performance. *Journal of Biological Education, 46,* 165-172. doi:10.1080/00219266.2011.617767

Long, D. (2012). Theories and models of student development. In L. J. Hinchliffe & M. A. Wong (Eds.), *Environments for student growth and development: Librarians and student affairs in collaboration* (pp. 41-55). Chicago: Association of College & Research Libraries.

Marks, G. N. (2017). Is SES really that important for educational outcomes in Australia? A review and some recent evidence. *The Australian Educational Researcher, 44,* 191-211. doi: 10.1007/s13384-016-0219-2.

Mattern, K., Shawn, E., & Williams, F. (2008). *Examining the relationship between the SAT®, high school measures of academic performance, and socioeconomic status: Turning our attention to the unit of analysis*. (College Board RN-36). Retrieved from: https://files.eric.ed.gov/fulltext/ED562606.pdf

McKenzie, K., & Schweitzer, R. (2001). Who succeeds at university? Factors predicting academic performance in first year Australian university students. *Higher Education Research and Development, 20*, 21-33. doi:10.1080/07924360120043621

*Morlaix, S., & Suchaut, B. (2014). The social, educational and cognitive factors of success in the first year of university: A case study. *International Review of Education, 60,* 841-862. doi:10.1007/s11159-014-9459-4

Mueller, C. W., & Parcel, T. L. (1981). Measures of socioeconomic status: Alternatives and recommendations. *Child Development, 52*, 13-30. doi:10.2307/1129211

Musso, M.F., & Cascallar, E.C. (2009). Predictive systems using artificial neural networks: An introduction to concepts and applications in education and social sciences. In M.C. Richaud & J.E. Moreno (Eds.), *Research in behavioral sciences* (Vol. 1, pp. 433-459). Buenos Aires, Argentina: CIIPME/CONICET.

Musso, M., Kyndt, E., Cascallar, E.C., & Dochy, F. (2012). Predicting mathematical performance: The effects of cognitive processes and self-regulation factors. *Education Research International*, *2012*, 1-13, doi:10.1155/2012/250719

Musso, M. F., Kyndt, E., Cascallar, E.C., & Dochy F. (2013). Predicting general academic performance and identifying the differential contribution of participating variables using artificial neural networks. *Frontline Learning Research*, *1*, 42-71. doi:10.14786/flr.v1i1.13

Nakao, K., & Treas, J. (1994). Updating occupational prestige and socioeconomic scores: How the new measures measure up. *Sociological Methodology*, *24*, 1-72. doi:10.2307/270978

National Institute for Health and Clinical Excellence (2009). *Methods for the development of NICE public health guidance* (2nd ed.). Retrieved from: http://www.nice.org.uk/media/2FB/53/PHMethodsManual110509.pdf

*Nguyen, T.M. (2016). Learning approaches, demographic factors to predict academic outcomes. *International Journal of Educational Management, 30,* 653-667. doi:10.1108/ijem-06-2014-0085

Nonis, S. A., & Hudson, G. I. (2006). Academic performance of college students: Influence of time spent studying and working. *Journal of Education for Business*, *81*, 151-159. doi:10.3200/joeb.81.3.151-159

*Pedrosa, R. H., Dachs, J. N. W., Maia, R. P., Andrade, C. Y., & Carvalho, B. S. (2007). Academic performance, students' background and affirmative action at a Brazilian University. *Higher Education Management and Policy*, *19*, 1-20. doi:10.1787/hemp-v19-art18-en

Peterson, R. A., & Brown, S. P. (2005). On the use of beta coefficients in meta-analysis. *Journal of Applied Psychology*, *90*, 175-181. doi: 10.1037/0021-9010.90.1.175

Pike, G. R., & Saupe, J. L. (2002). Does high school matter? An analysis of three methods of predicting first-year grades. *Research in Higher Education*, *43*, 187-207.

Pitman, T., Trinidad, S., Devlin, M., Harvey, A., Brett, M., & McKay, J. (2016). *Pathways to higher education: The efficacy of enabling and sub-bachelor pathways for disadvantaged students*. National Centre for Student Equity in Higher Education (NCSEHE), Perth: Curtin University.

*+Puddey, I. B., & Mercer, A. (2014). Predicting academic outcomes in an Australian graduate entry medical programme. *BMC Medical Education*, *14*, 1-12. doi:10.1186/1472-6920-14-31

Reason, R. D. (2009). An examination of persistence research through the lens of a comprehensive conceptual framework. *Journal of College Student Development*, *50*, 659-682. doi:10.1353/csd.0.0098

Richardson, M., Abraham, C., & Bond, R. (2012). Psychological correlates of university students' academic performance: A systematic review and meta-analysis. *Psychological Bulletin*, *138*, 353-387. doi:10.1037/a0026838

Roberts, A. S. (2007). Predictors of future performance in architectural design education. *Educational Psychology*, *27*, 447-463. doi:10.1080/01443410601104361

Rochford, C., Connolly, M., & Drennan, J. (2009). Paid part-time employment and academic performance of undergraduate nursing students. *Nurse Education Today, 29*, 601-606. doi:10.1016/j.nedt.2009.01.004

*Rodríguez Albor, G., Ariza Dau, M., & Ramos Ruíz, J. L. (2014). Calidad institucional y rendimiento académico: El caso de las universidades del Caribe colombiano. [Institutional quality and academic performance: The case of Colombian Caribbean universities]. *Perfiles Educativos, 36,* 10-29. doi:10.1016/s0185-2698(14)70607-5

*+Rodríguez Ayan, M.N., & Ruiz Díaz, M. A. (2011). Indicadores de rendimiento de estudiantes universitarios: Calificaciones versus créditos acumulados. [Performance's indicators of college students: Grades versus accumulated credits]. *Revista de Educación, 355,* 467-492. doi:10-4438/1988-592X-RE-2011-355-033

Rothstein, H. R., Sutton, A. J., & Borenstein, M. (2005). *Publication bias in meta-analysis: Prevention, assessment and adjustments*. Chichester, UK: John Wiley.

*+Rothstein, J. M. (2004). College performance predictions and the SAT. *Journal of Econometrics, 121*, 297-317. doi:10.1016/j.jeconom.2003.10.003

*+Sackett, P. R., Kuncel, N. R., Arneson, J. J., Cooper, S. R., & Waters, S. D. (2009). Does socioeconomic status explain the relationship between admissions tests and post-secondary academic performance? *Psychological Bulletin, 135*, 1-22. doi:10.1037/a0013978

Sackett, P. R., Kuncel, N. R., Beatty, A. S., Rigdon, J. L., Shen, W., & Kiger, T. B. (2012). The role of socioeconomic status in SAT-grade relationships and in college admissions decisions. *Psychological Science, 23,* 1000-1007. doi:10.1177/0956797612438732

Sadler, D. R. (2013). Making competent judgments of competence. In Blömeke S., Zlatkin-Troitschanskaia O., Kuhn C., Fege J. (Eds.), *Professional and Vet Learning: Vol. 1. Modeling and Measuring Competencies in Higher Education* (pp. 13-27). Dordrecht, The Netherlands: Sense Publishers. doi:10.1007/978-94-6091-867-4

Scheerens, J. (1990). School effectiveness research and the development of process indicators of school functioning. *School Effectiveness and School Improvement*, *1*, 61-80. doi: 10.1080/0924345900010106

Schneider, M., & Preckel, F. (2017). Variables associated with achievement in higher education: A systematic review of meta-analyses. *Psychological Bulletin, 143,* 565-600. doi:10.1037/bul0000098

Schneider, S. L. (2013). The international standard classification of education 2011. In G.E. Birkelund (Ed.). *Comparative Social Research: Vol. 30. Class and stratification analysis* (pp. 365-379). Bingley, UK: Emerald Group Publishing Limited. doi: 10.1108/s0195-6310(2013)0000030017

Shavelson, R. J. (2010). On the measurement of competency. *Empirical Research in Vocational Education and Training*, *2*, 41-63.

Shavers, V. L. (2007). Measurement of socioeconomic status in health disparities research. *Journal of the National Medical Association*, *99*, 1013-1023.

*Shulruf, B., Hattie, J., & Tumen, S. (2008). Individual and school factors affecting students' participation and success in higher education. *Higher Education, 56,* 613-632. doi:10.1007/s10734-008-9114-8

Sidhu, K. S. (2005). *New approaches to measurement and evaluation*. New Delhi: India. Sterling Publishers Private Limited.

Sirin, S. (2005). Socioeconomic status and academic achievement: A meta-analytic review of research. *Review of Educational Research, 75*, 417-453. doi:10.3102/00346543075003417

*+Smith, E. (2016). Can higher education compensate for society? Modelling the determinants of academic success at university. *British Journal of Sociology of Education, 37,* 970-992. doi:10.1080/01425692.2014.987728

Stinebrickner, R., & Stinebrickner, T. R. (2003). Working during school and academic performance. *Journal of Labor Economics, 21*, 473-491. doi:10.1086/345565

*Stratton, L. S., & Wetzel, J. N. (2011, May). *The role of socioeconomic status when controlling for academic background in a multinomial logit model of six-year college outcomes*. Paper presented at the 51st Annual Forum of the Association for Institutional Research, Toronto, ON.

Sullivan, A. (2001). Cultural capital and educational attainment. *Sociology, 35*,893-912. doi:10.1017/s0038038501008938

*+Tai, R. H., Sadler, P. M., & Loehr, J. F. (2005). Factors influencing success in introductory college chemistry. *Journal of Research in Science Teaching, 42,* 987-1012. doi:10.1002/tea.20082

The World Bank (2012). *Reviews of national policies for education: Tertiary education in Colombia 2012*. Paris: OECD Publishing. doi:10.1787/9789264180697-en

*+Thiele, T., Singleton, A., Pope, D., & Stanistreet, D. (2016). Predicting students' academic performance based on school and socio-demographic characteristics. *Studies in Higher Education, 41*, 1424-1446. doi:10.1080/03075079.2014.974528

Tinto, V. (1975). Dropout from higher education: A theoretical synthesis of recent research. *Review of Educational Research, 45*, 89-125.

Tinto, V. (1993). *Leaving college. Rethinking the causes and cures of student attrition.* Chicago, IL: The University of Chicago Press.

Tracey, T. J., Allen, J., & Robbins, S. B. (2012). Moderation of the relation between person–environment congruence and academic success: Environmental constraint, personal flexibility and method. *Journal of Vocational Behavior*, *80*, 38-49. doi:10.1016/j.jvb.2011.03.005

*Triventi, M. (2014). Does working during higher education affect students' academic progression? *Economics of Education Review, 41*, 1-13. doi:10.1016/j.econedurev.2014.03.006

Van Ewijk, R., & Sleegers, P. (2010). The effect of peer socioeconomic status on student achievement: A meta-analysis. *Educational Research Review, 5*, 134-150. doi:10.1016/j.edurev.2010.02.001

*+Walpole, M. (2003). Socioeconomic status and college: How SES affects college experiences and outcomes. *The Review of Higher Education, 27*, 45-73. doi:10.1353/rhe.2003.0044

*+Wang, H., Kong, M., Shan, W., & Vong, S. K. (2010). The effects of doing parttime jobs on college student academic performance and social life in a Chinese society. *Journal of Education and Work, 23*, 79-94. doi:10.1080/13639080903418402

*+Waqas, M., Abbasi, A., & Idrees, R. (2013). Academic performance of undergraduates in universities of Pakistan. *Middle-East Journal of Scientific Research, 16*, 55-61.

Warburton, E., Bugarin, R., & Nuñez, A. (2001). *Bridging the gap: Academic preparation and postsecondary success of first-generation students* (NCES 2001–153). Washington, DC: National Center for Education Statistics, U.S. Government Printing Office.

Westrick, P. A., Le, H., Robbins, S. B., Radunzel, J. M., & Schmidt, F. L. (2015). College performance and retention: A meta-analysis of the predictive validities of ACT® scores, high school grades, and SES. *Educational Assessment, 20*, 23-45. doi:10.1080/10627197.2015.997614

White, K. (1982). The relation between socioeconomic status and academic achievement. *Psychological Bulletin, 91*, 461-481. doi:10.1037//0033-2909.91.3.461

*+Win, R., & Miller, P. W. (2005). The effects of individual and school factors on university students' academic performance. *Australian Economic Review, 38*, 1-18. doi:10.1111/j.1467-8462.2005.00349.x

*+Wolniak, G. C., & Engberg, M. E. (2010). Academic achievement in the first year of college: Evidence of the pervasive effects of the high school context. *Research in Higher Education, 51*, 451-467. doi:10.1007/s11162-010-9165-4

*+Yanbarisova, D. M. (2015). The effects of student employment on academic performance in Tatarstan higher education institutions. *Russian Education & Society, 57*, 459-482. doi:10.1080/10609393.2015.1096138

*+Yao, G., Zhimin, L., & Peng, F. (2015). The effect of family capital on the academic performance of college students. A survey at 20 higher education institutions in Jiangsu province. *Chinese Education & Society, 48*, 81-91. doi:10.1080/10611932.2015.1014713

York, T. T., Gibson, C., & Rankin, S. (2015). Defining and measuring academic success. *Practical Assessment, Research & Evaluation, 20*, 1-20.

Yusuf, A. (2002). Interrelationships among academic performance, academic achievement and learning outcomes. *Journal of Curriculum and Instruction, 1*, 87-96.

Zlatkin-Troitschanskaia, O., Shavelson, R. J., & Kuhn, C. (2015). The international state of research on measurement of competency in higher education. *Studies in Higher Education, 40*, 393-411. doi:10.1080/03075079.2015.1004241

*Zheng, J. L., Saunders, K. P., Shelley, I. I., Mack, C., & Whalen, D. F. (2002). Predictors of academic success for freshmen residence hall students. *Journal of College Student Development, 43*, 267-283.

Zwick, R. (2012). The role of admissions test scores, socioeconomic status, and high school grades in predicting college achievement. *Pensamiento Educativo. Revista de Investigación Educacional Latinoamericana, 2,* 23-30. doi:10.7764/pel.49.2.2012.3

Table 1

Overview of the literature search hits.

Query	Search terms	ERIC (Ovid)	PsycArticles	Scopus	SSCI	
1	¨Higher Education¨ AND ¨Academic Achievement¨	13,668	429	988	345	*15,430*
	Update May 2017	333	17	110	80	*540*
2	¨Higher Education¨ AND ¨Academic Performance¨	9,535	287	752	553	*11,127*
	Update May 2017	264	9	186	145	*604*
3	¨Higher Education¨ AND ¨Academic Outcomes¨	0	1708	83	66	*1,857*
	Update May 2017	395[2]	78	20	15	*508*
Sub-total initial search		23,203	2424	1823	964	*28,414*
Sub-total update May 2017		992	104	316	240	*1,652*
Total		**24,195**	**2,528**	**2,139**	**1,204**	***30,066***

[2] This number of hits was obtained using the multi-field function and selecting references between 2000 and 2017.

Table 2

Operationalizations of SES.

Measurement	Type of measurement	Scale	Operationalization	Reference(s)
Parental Educational level (n=25)	As a single indicator (n=21)	Categorical (n=19)	Highest level of parental education (n=4)	Bruinsma and Jansen (2007); Ifenthaler and Widanapathirana (2014); Loehr et al. (2012); Tai et al. (2005)
			A dichotomous variable to determine whether parents have attended college (n=3)	Bonsaksen (2016); Nguyen (2016); Waqas et al. (2013)
			3 categories	Rodríguez and Ruiz (2011)
			4 categories	Stratton and Wetzel (2011)
			4 categories	De Clercq et al. (2013)
			4 categories (years of education)	Triventi (2014)
			5 categories	Beyene and Yimam (2016)
			5 categories	Frischenschlager et al. (2005)
			5 categories	Gerken and Volkwein (2000)
			5 categories	Harb and El-Shaarawi (2007)
			7 categories	Guimarães and Sampaio (2013)
			Only father's educational level (8 categories)	Morlaix and Suchaut (2014)
			Only mother's educational level (3 categories)	Wolniak and Engberg (2010)
			Only mother's educational level (5 categories)	Black et al. (2015)
		Continuous (n=2)	Years of education (average)	Rothstein (2004)
			Years of education: going from 8 to 19	Delaney et al. (2011)
	Within a composite score (n=4)	Categorical (n=3)	Highest level of parental education (n=2)	De Clercq et al. (2017); Rodríguez et. al (2014)
			ISCED (International Standard Classification of Education)	Gouvias et al. (2012)
		Continuous	Occupational prestige and socioeconomic scores (Nakao & Treas, 1994).	Walpole (2003)

Table 2 (continued)

Measurement	Type of measurement	Scale	Operationalization	Reference(s)
Parental Occupation (n=10)	As a single indicator (n=7)	Categorical (n=7)	Dichotomous variable to indicate whether parents work	Guimarães and Sampaio (2013)
			Dichotomous variable to indicate whether parents' occupation(s) requires higher education	Bonsaksen (2016)
			4 categories	Yao et al. (2015)
			5 categories	Arulampalam et al. (2004)
			8 categories	Morlaix and Suchaut (2014)
			10 categories	Hansen and Mastekaasa (2006)
			The Standard Occupational Classification scale 2000	Smith (2016)
	Within a composite score (n=3)	Categorical (n=2)	7 categories	Gouvias et al. (2012)
			12 categories	Rodríguez et al. (2014)
		Continuous	Occupational prestige and socioeconomic scores (Nakao & Treas, 1994).	Walpole (2003)
Income (n=20)	As a single indicator (n= 17)	Categorical (n=16)	3 categories (high, middle and low) (n=3)	Frischenschlager et al. (2005); Wolniak and Engberg (2010); Yao et al. (2015)
			5 categories	Black et al. (2015)
			5 categories	Stratton and Wetzel (2011)
			5 categories (quintiles)	Bozick (2007)
			6 categories	Yanbarisova (2015)
			7 categories	Guimarães and Sampaio (2013)
			7 categories	Waqas et al. (2013)
			Financial aid (receiving a scholarship or a loan) (n=4)	Black et al. (2015); Morlaix and Suchaut (2014); Stratton and Wetzel (2011); Triventi (2014)
			To receive subsidized lunches (n=2)	Black et al. (2015); Rothstein (2004)
			University tuition fees (quartiles)	Triventi (2014)
		Continuous	Parents' relative income going from 0 to 1	Hansen and Mastekaasa (2006)
	Within a composite score (n=3)	Continuous (n=3)	Monthly income	Rodríguez et. al (2014)
			Occupational prestige and socioeconomic scores (Nakao & Treas, 1994)	Walpole (2003)
			University tuition fees	Rodríguez et al. (2014)

Table 2 (continued)

Measurement	Type of measurement	Scale	Operationalization	Reference(s)
Household resources (n=8)	As a single indicator (n=6)	Categorical (n=4)	Computer at home and internet connection (n=3)	Gouvias et. al (2012); Guimarães and Sampaio (2013); Pedrosa et al. (2007)
			Possession of books related to schoolwork	Gouvias et al. (2012)
		Continuous (n=2)	Household crowding (n=2)	Harb and El-Shaarawi (2007); Gouvias et. al (2012)
	Within a composite score (n=2)	Continuous (n=2)	Number of possessions (car, books, musical instruments, computer) (n=2)	De Clercq et al. (2017), Pedrosa et al. (2007)
Neighborhood resources (n=9)	As a single indicator (n=3)	Categorical (n=3)	Students' stratum (n=2)	Bahamón and Reyes Ruiz (2014); Rodríguez et al. (2014)
			Degree of urbanization	Hansen and Mastekaasa (2006)
	As a composite score (n=6)	Continuous (n=6)	Index of Economic Resources (IER) (n=2)	Birch and Miller (2006); Win and Miller (2005)
			Index of Education and Occupation (IEO)	Win and Miller (2005)
			Index of Multiple Deprivation (IMD)	Thiele et al. (2016)
			Index of Relative Socioeconomic Advantage and Disadvantage (IRSAD)	Puddey and Mercer (2014)
			School decile	Shulruf et al. (2008)

Table 3a

Operationalizations of academic performance at university.

Measurement	Operationalization		Type of scale	Scale	Reference(s)
Achievement (n=43)	GPA (n=32)	First year GPA (n=16)	Categorical (n=4)	3 categories (high, medium, low) (n=2)	Shulruf et al. (2008); Yao et al. (2015)
				4 categories	De Clercq et al. (2017)
				5 categories	Hansen and Mastekaasa (2006)
			Continuous (n=12)	From 0 to 4 (n=6)	Anderton et al. (2016); Black et al. (2015); Rothstein (2004); Sackett et al. (2009); Wolniak and Engberg (2010); Zheng et al. (2002)
				From 1 to 10	Bruinsma and Jansen (2007)
				From 0 to 20	Morlaix and Suchaut (2014)
				From 0 to 100 (n=4)	Birch and Miller (2006); De Clercq et al. (2013); De Clercq et al. (2017); Win and Miller (2005)
		GPA (n=8)	Categorical (n=2)	3 categories	Yao et al. (2015)
				3 categories	Yanbarisova (2015)
			Continuous (n=6)	From 0 to 4 (n=5)	Black et al. (2015); Bonsaksen (2016); Harb and El-Shaarawi (2007); Sackett et al. (2009); Wang et al. (2010)
				From 0 to 12	Rodríguez and Ruiz (2011)
		Final GPA (n=8)	Categorical (n=2)	3 categories	Yao et al. (2015)
				5 categories	Waqas et al. (2013)
			Continuous (n=6)	From 0 to 4 (n=4)	Black et al. (2015); Gerken and Volkwein (2000); Sackett et al. (2009); Zheng et al. (2002)
				From 0 to 7	Puddey and Mercer (2014)
				From 0 to 100	Thiele et al. (2016)
	Single grade (n=4)	Test at the end of the first year	Categorical	High or low performance	Frischenschlager et al. (2005)
		Introductory college chemistry	Continuous (n=3)	From 0 to 12	Rodríguez and Ruiz (2011)
		Introductory college chemistry		From 0 to 100	Tai et al. (2005)
		Introductory college biology		From 0 to 100	Loehr et al. (2012)

Table 3a (continued)

Operationalizations of academic performance at university.

Measurement	Operationalization		Type of scale	Scale	Reference(s)
	Number of credits (n=6)	Course credits (n=2)	Continuous (n=4)		Black et al. (2015); Gerken and Volkwein (2000)
		Study unit outcomes (n=2)			Ifenthaler and Widanapathirana (2014); Puddey and Mercer (2014)
		Degree progression index	Continuous	From 0 to 1	Rodríguez and Ruiz (2011)
		ECTS credits	Continuous		Triventi (2014)
	Relative achievement	Index to compare students' performance at the entrance and exit of all courses	Continuous	From -1 to 1	Pedrosa et al. (2007)
Competencies (n=6)	Standardized test (n=6)	SABER PRO (Colombia) (n=2)	Continuous (n=2)	From 1 to 300	Bahamón and Reyes Ruiz (2014); Rodríguez et al. (2014)
		CET (College English Test) (China)	Continuous	The highest possible score is 710	Yao et al. (2015)
		DALF (France)	Continuous	From 0 to 100	Morlaix and Suchaut (2014)
		GAMSAT (Australia)	Continuous	From 0 to 100	Puddey and Mercer (2014)
		PELA (Australia)	Continuous	From 0 to 10	Anderton et al. (2016)
Persistence (n=6)	Completion status (n=6)	Dropout (n=2)	Categorical (n=2)	Yes/No	Arulampalam et al. (2004); Gerken and Volkwein (2000)
		Final degree classification (n=2)	Categorical (n=2)	British undergraduate degree classification	Smith (2016); Thiele et al. (2016)
		Attending graduate school	Categorical	Yes/No	Walpole (2003)
		Graduation rate	Continuous		Stratton and Wetzel (2011)

Table 3b

Operationalizations of academic performance before university.

Measurement	Operationalization		Type of scale	Scale	Reference
Achievement (n=20)	HSGPA (n=14)		Categorical (n=4)	2 categories	Smith (2016)
				4 categories	Stratton and Wetzel (2011)
				4 categories	Triventi (2014)
				4 categories	Morlaix and Suchaut (2014)
			Continuous (n=10)	From 0 to 4 (n=4)	Gerken and Volkwein (2000); Rothstein (2004); Wolniak and Engberg (2010); Zheng et al. (2002)
				From 1 to 4	De Clercq et al. (2017)
				From 1 to 5 (being 1 the best, and 5 the worst)	Frischenschlager et al. (2005)
				From 1 to 6	Hansen and Mastekaasa (2006)
				From 1 to 10	Bruisnma and Jansen (2007)
				From 1 to 20	Gouvias et al. (2012)
				UCAS tariff points	Thiele et al. (2016)
	Single grade (n=3)		Continuous (n=3)	HS grades in science, mathematics, and English courses (n=2)	Loerh et al. (2012); Tai et al. (2005)
				HS grades in biology, chemistry and physics courses	Arulampalam et al. (2004)
	Subject area exams (n=3)	Leaving Certificate (Ireland)	Continuous	From 0 to 600	Delaney et al. (2011)
		NCEA (New Zealand)	Continuous	From 0 to 80 credits (each NCEA level)	Shulruf et al. (2008)
		SABER 11 (Colombia)	Continuous	Before 2014: From 0 to 400 As of 2014: From 0 to 500	Bahamón and Reyes Ruiz (2014)

Table 3b (continued)

Operationalizations of academic performance before university

Measurement	Operationalization		Type of scale	Scale	Reference
Competencies (n=15)	Aptitude exams (n=12)	SAT (USA) (n=9)	Continuous (n=9)	Before 2016: from 600 to 2400 As of 2016: from 400 to 1600	Black et. al. (2015); Gerken and Volkwein (2000); Loehr et al. (2012); Rothstein (2004); Sackett et.al. (2009); Stratton and Wetzel (2011); Tai et al. (2005); Walpole (2003); Wolniak and Engberg (2010)
		ACT (USA) (n=3)	Continuous (n=3)	From 1 to 36	Stratton and Wetzel (2011); Wolniak and Engberg (2010); Zheng et al. (2002)
	Subject area exams (n=3)	Vestibular (Brazil)	Continuous (n=3)		Guimarães and Sampaio (2013)
		Not specified (Greece)			Gouvias et al. (2012)
		Not specified (Vietnam)			Nguyen (2016)
Persistence (n=8)	Admission ranks (n=5)	High school rank (n=2)	Continuous (n=2)		Black et al. (2015); Zheng et al. (2002)
		Tertiary Entrance Rank (TER) (n=2)	Continuous (n=3)	From 0 to 99.95	Birch and Miller (2006); Win and Miller (2005)
		Australian Tertiary Admission Rank (ATAR)			Anderton et al. (2016)
	Grade retention (n=3)	Years repeated in high school	Categorical	Yes/No	De Clercq et al. (2013)
		Number of class repetitions	Continuous (n=2)		Frischenschlager et al. (2005)
		Years repeated in high school			Morlaix and Suchaut (2014)

Table 4

The relationship between SES and academic performance in higher education.

Reference	N	SES-Academic performance relationship	Average effect size	Prior academic achievement (PAA) measure	PAA-AP relationship	Average effect size	University experience (UE) measure	UE-AP relationship	Average effect size	Working status (WS) measure	WS-AP relationship	Average effect size
Anderton et al. (2016)	414	Not significant		ATAR score	Positive							
Arulampalam et al. (2004)	51,810	Negative	-0.01	HSGPA	Negative	-0.02						
Bahamón and Reyes Ruiz (2014)	68	Positive										
Beyene and Yimam (2016)	925	Positive	0.14	Entrance exam score	Positive	0.05	Academic experience (number of assessments)	Positive	0.12			
Birch and Miller (2006)	1,803	Negative	-0.04^{+}	TER score	Positive	0.58						
Black et al. (2015)	23,792	Negative	-0.32^{+}									
Bonsaksen (2016)	123	Not significant	0.02	Prior experience in higher education	Positive	0.26				Employed or not	Not significant	-0.04
Bozick (2007)	10,164	Positive	0.01							Employed or not	Negative	-0.13
Bruinsma and Jansen (2007)	62	Not significant	0.06	HSGPA	Positive	0.56	Academic experience (classroom climate, quantity and quality of the instruction)	Negative	-0.04	Employed or not	Not significant	0.07
De Clercq et al. (2013)	111	Positive	0.26	Failure in secondary school	Negative	-0.21						
De Clercq et al. (2013)	206	Not significant	0.15	Failure in secondary school	Negative	-0.19	Social experience (peer support)	Not significant	0.15			
De Clercq et al. (2017)	2178	Positive	0.12	HSGPA	Positive	0.37						
Delaney et al. (2011)	1,867	Not significant	-0.02	Entrance exam score	Positive	0.31						
Frischenschlager et al. (2005)	245	Not significant		HSGPA	Positive							

Gerken and Volkwein (2000)	-	Not significant		HSGPA	Positive		Academic experience	Positive					
Gouvias et al. (2012)	874	Positive	0.06	HSGPA	Positive	0.79							
Guimarães and Sampaio (2013)	54,877	Positive	0.09										
Hansen and Mastekaasa (2006)	56,792	Positive	0.02	HSGPA	Positive	0.36							
Harb and El-Shaarawi (2007)	296	Negative	-0.05	Science HS diploma	Positive	0.11	Institutional experience (positive attitude towards university)	Positive	0.17	Employed or not	Negative	-0.3	
Ifenthaler and Widanapathirana (2014)	146,001	Not significant											
Loehr et al. (2012)	2,667	Positive		HSGPA	Positive								
Morlaix and Suchaut (2014)	543	Negative		HSGPA, no failure in secondary school	Positive								
Nguyen (2016)	616	Not significant											
Pedrosa et al. (2007)	6,701	Negative		HS school courses	Positive								
Puddey and Mercer (2014)	219	Positive	0.17^{+}	GPA at entry of graduate program, GAMSAT score	Positive	0.25							
Rodríguez et al. (2014)	14,829	Positive								Employed or not	Negative		
Rodríguez and Ruiz (2011)	312	Not significant	-0.02							Employed or not	Negative	-0.22	
Rothstein (2004)	14,102	Positive	0.07	HSGPA, SAT	Positive	0.71							
Sackett et al. (2009)	17,630	Positive	0.09	SAT	Positive (n=17,244)	0.37							
Shulruf et al. (2008)	1,880	Positive	0.01^{+}										
Smith (2016)	23,793	Positive	0.03	HSGPA	Positive	0.34							
Stratton and Wetzel (2011)	5,823	Positive		HSGPA	Positive								
Tai et al. (2005)	1,333	Positive	0.12	HSGPA, SAT	Positive	0.17							

Reference	N	SES-Academic performance relationship	Average effect size	Prior academic achievement (PAA) measure	PAA-AP relationship	Average effect size	University experience (UE) measure	UE-AP relationship	Average effect size	Working status (WS) measure	WS-AP relationship	Average effect size
Thiele et al. (2016)	3,730	Positive	0.05+	UCAS points	Positive	0.003						
Triventi (2014)	1,834	Positive										
Walpole (2003)	6,470	Positive	0.01	GPA at entry of graduate program	Positive (n=1177)	0.09	Academic experience (working on research)	Positive (n=1177)	0.13			
Wang et al. (2010)	323	Positive	0.06				Institutional experience (sense of belonging and school integration)	Positive	0.12	Reasons for working and characteristics of work	Positive	0.06
Waqas et al. (2013)	267	Positive	0.05							Employed or not	Negative	-0.14
Win and Miller (2005)	1,803	Positive	0.03+	TER score	Positive	0.1						
Wolniak and Engberg (2010)	3,750	Positive		HSGPA, SAT	Positive	0.26						
Yanbarisova (2015)	1,988	Not significant	0.03							Characteristics of work	Negative	-0.01
Yao et al. (2015)	2,989	Positive	0.16									
Zheng et al. (2002)	1,166	Not significant		HSGPA	Positive							

(+) Effect sizes calculated from SES indicators at area level. Not considered in the analysis.

Figure 1

Conceptual diagram for the analysis of the mediators.

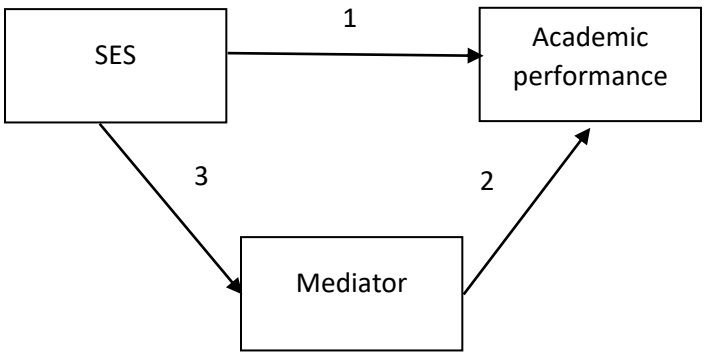


Figure 2

Authors' model for measuring SES.

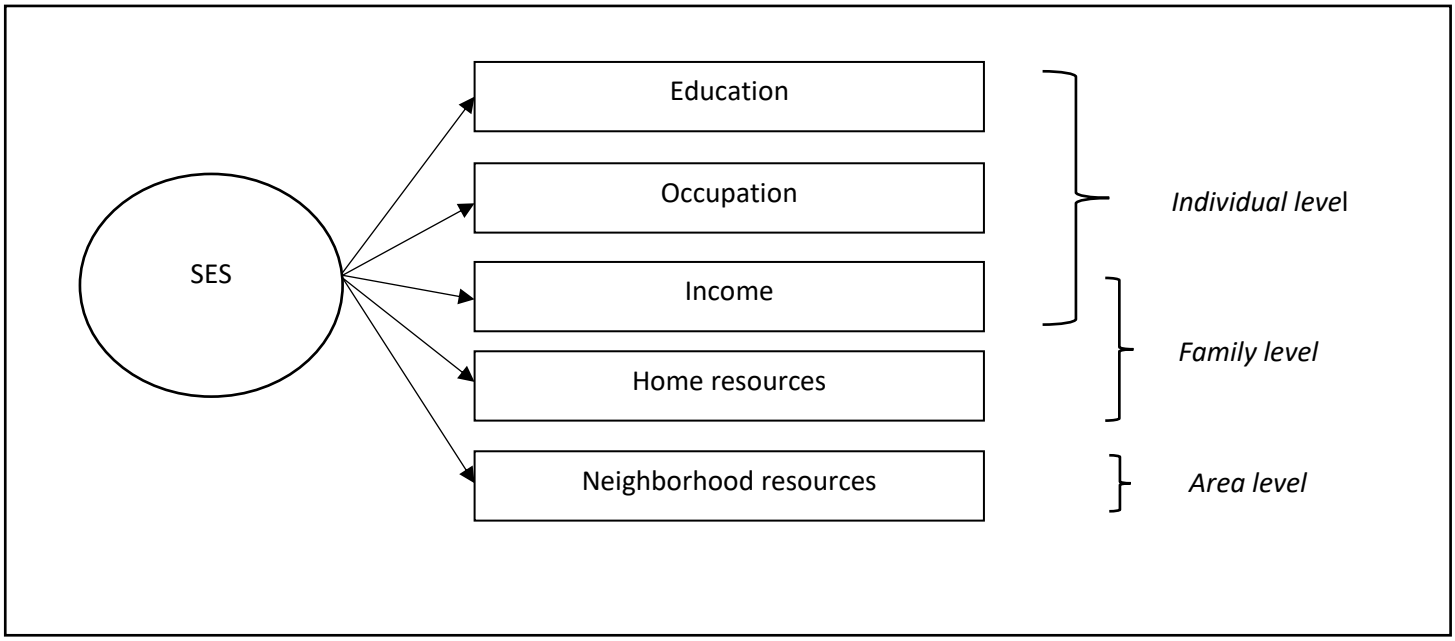


Figure 3

Authors' model for measuring academic performance in higher education.

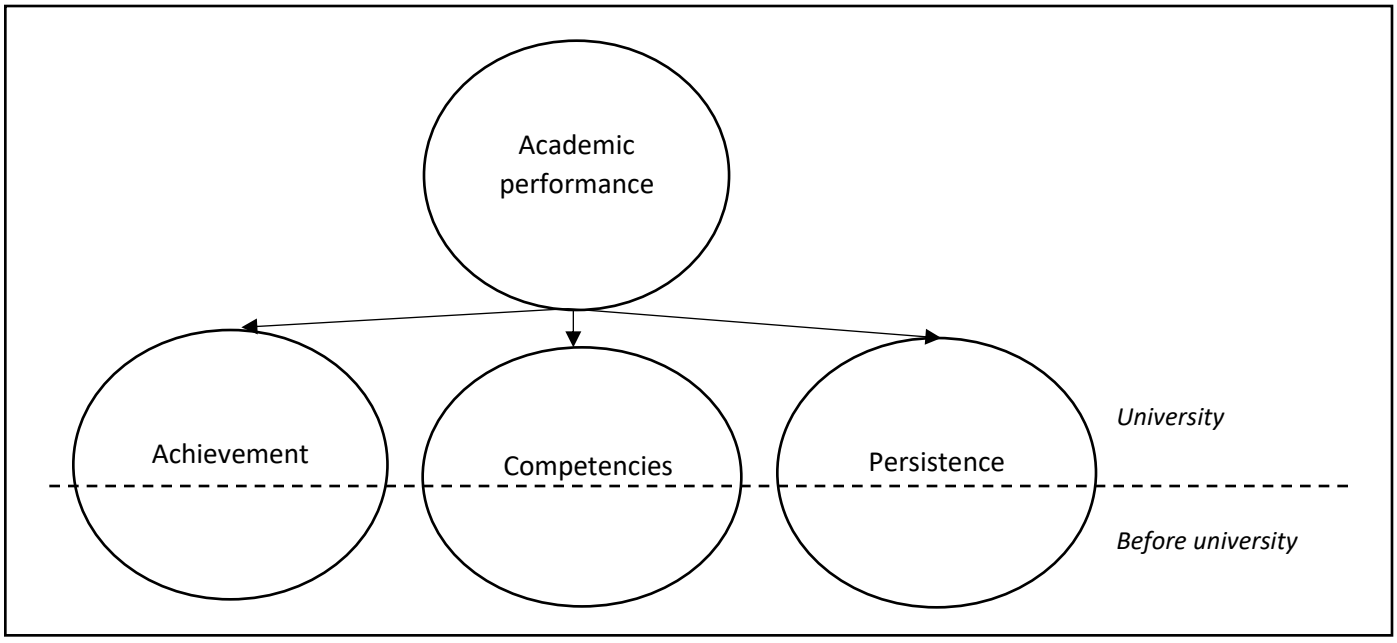
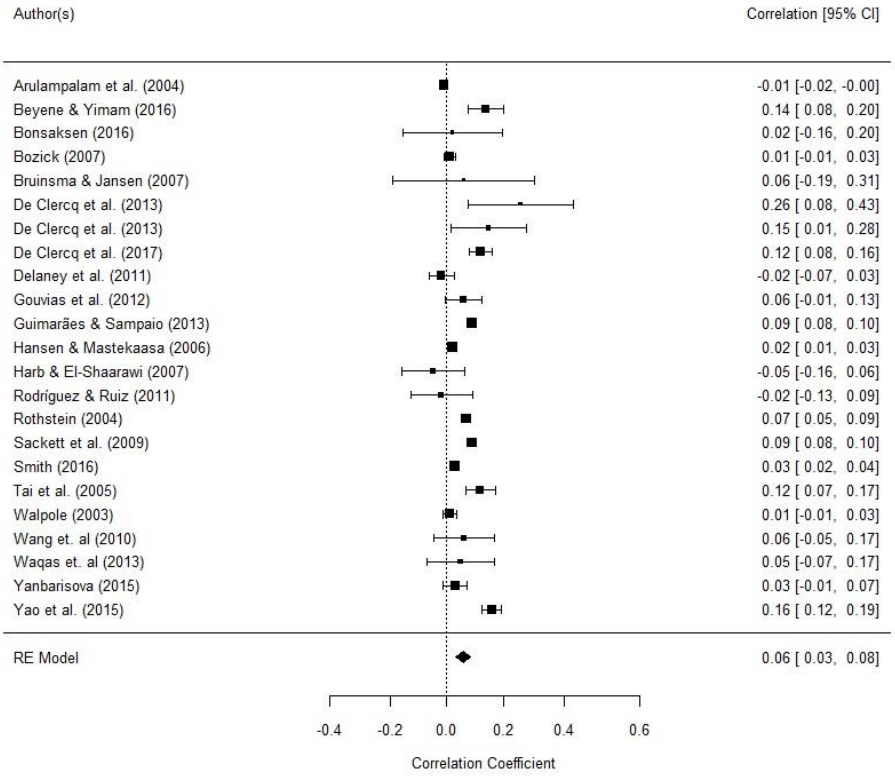


Figure 4

Forest plot of the relationship between SES and academic performance in higher education.



Chapter 3

Multistage approach for using artificial neural networks to classify academic performance in higher education[3]

Abstract

Several limitations have prevented the use of artificial neural networks (ANNs) in educational research. In particular, the absence of standard design rules, the lack of understanding of their quality measures, and the misconception that it is not possible to interpret ANNs models persist among educational scientists. To address such limitations, the present study proposes a multistage approach for the use of ANNs to classify the academic performance of 162,030 university students. Three stages are followed, namely, design, evaluation, and interpretation. Findings suggest that it is possible to systematically train and test ANNs models to classify students' academic performance as either high-performance (accuracy of 82%) or low-performance (accuracy of 75%); to assess the quality of the derived predictions using several quality measures; and to provide clear and transparent interpretations of the predictive models based on ANNs. Furthermore, prior academic achievement, socio-economic conditions, and high-school characteristics continue to be weighty predictors of students' academic performance in higher education.

Keywords: artificial neural networks, academic performance, higher education, predictive systems.

[3] The text in this chapter is currently under review by the *International Journal of Artificial Intelligence in Education* as Rodríguez-Hernández, C.F., Musso, M., Kyndt, E., Cascallar, E. Multistage approach for using artificial neural networks to classify academic performance in higher education.

1. Introduction

The prediction of academic performance in higher education provides several benefits to teachers, students, policymakers, and institutions. For instance, to improve the selection process of students who should receive grants or scholarships, to avoid future academic failure and dropout, and to identify which teaching practices could have a more positive impact on students' learning (Musso, Kyndt, Cascallar, & Dochy, 2013). Although first-year grade point average (FYGPA) is commonly used to assess academic performance in higher education (Noble & Sawyer, 2002), relatively more attention has been given to standardized tests as an assessment tool of this construct.

The current body of literature has focused mainly on predicting students' academic performance using traditional statistical techniques (e.g., Ayán & García, 2008; Huang & Fang, 2010). Golino and Gomes (2014) report that correlation coefficients, linear and logistic regression analysis, analysis of variance (ANOVA), and structural equation modeling (SEM) are the most common methods when predicting academic performance. However, these traditional statistical techniques need to fulfill a set of assumptions such as independence of observations, homogeneity of variance, normality, and linearity (Tacq, 1997; Field, 2009). Remarkably, several of these assumptions are not always reported by researchers. Potential undesirable effects of violating statistical assumptions such as type I and II errors (Nimon, 2012) and inadequate estimation of the effect sizes of statistical parameters (Osborne & Waters, 2002) are ignored. Therefore, in some cases, predictive models based on traditional statistical techniques could be methodologically biased.

Furthermore, higher education faces the challenge of collecting, storing, and analyzing information related to students' academic performance in a non-traditional way (Daniel, 2017).

When predicting students' academic performance in large data sets using traditional statistical techniques, inaccurate interpretations and conclusions might be drawn. As the sample size of the data set increases (as in large-scale testing outcomes) p-values approaches to zero, so results are always statistically significant due to the sample size rather than the relationship actually being meaningful (Khalilzadeh & Tasci, 2017).

Starting with the drawbacks of the prediction of the students' academic performance using traditional statistical techniques, recent educational research has used predictive models based on artificial neural networks (ANNs) (e.g., Musso et al., 2013; Kyndt, Musso, Cascallar, & Dochy, 2015). The use of ANNs as predictive systems of students' academic performance has shown several advantages over traditional statistical techniques. Predictive systems entail the possibility of using all the interactions between predictor variables to achieve a better estimation of the outcome variable (Cascallar, Musso, Kyndt, & Dochy, 2015) and possess the capability of obtaining a prediction even when the independent and dependent variables are related in a non-linear way (Somers & Casal, 2009). In addition, ANNs allow the analysis of vast volumes of information and to build predictive models regardless of the statistical distribution of the data (Garson, 2014). However, up till now, the use of ANNs in the field of education has not received much attention.

Some limitations have been pointed out in the literature regarding the use of ANNs when predicting students' academic performance. First, it has been argued that there are no specific rules for designing ANNs, conceiving this process as a trial and error task. Second, there is still a limited understanding of how the quality of the predictions resulting from models can be assessed. Finally, ANNs have been commonly understood as "black boxes", where it is not possible to interpret the importance of the predictors when estimating an output variable.

This study will address the limitations presented above by proposing a multistage approach for using ANNs to predict academic performance in higher education. As such, three stages will be discussed: *design, evaluation, and interpretation*. To give an illustration of this multistage approach, the outcomes in the SABER PRO test (a standardized test taken in Colombia at the end of the university studies) in a cohort of 162,030 students will be predicted. Prior academic achievement, socio-economic status, high school characteristics, and working status will be selected as inputs of the predictive systems based on ANNs, and it is expected that the predictive weight of each input will be identified. All things considered, the main purpose of this study is to contribute to the methodology used in the application of these machine-learning procedures in educational research.

2. State of the art

This section contains information regarding artificial neural networks (ANNs), predictors of academic performance in higher education, and standardized tests as an indicator of academic performance in higher education.

2.1. Artificial Neural Networks: A Short Introduction

An artificial neural network is an information processing system formed by processing units called neurons (Garson, 2014; Haykin, 2009). A neuron is the most fundamental element in any artificial neural network, which allows mapping different inputs onto an output. To do so, each neuron has a sum block where the inputs are summed after being weighted; and an activation function, where the output is calculated after receiving the result from the sum block. The weights represent the strength of the connection between the neurons and the information used by the network to minimize the error with respect to the output (Fausett, 1994). In general, any artificial neural network could be characterized by (1) its arrangement of connections between the neurons

(called its topology), (2) its approach to obtaining the strength or the weights on the connections (called its training or learning algorithm), and (3) its activation function (Fausett, 1994).

The Multilayer Perceptron (MLP) is one widely known topology of ANNs. The MLP has a structure with at least three layers: the input layer, the hidden layer, and the output layer. As in the traditional statistical techniques, the input layer represents the independent variables or predictors; the hidden layer is where the mapping to relate input and output takes place, and the output layer resembles the dependent variable (Somers & Casal, 2009). As such, MLPs can be used to solve both prediction and classification problems (Garson, 2014; Haykin, 2009).

One common way to train the MLP is "supervised learning". In classification problems, the purpose of the training is to determine whether the subjects belong to specific groups or categories based on the set of predictors. As such, there is a target category for which classification accuracy is maximized. As there is also information on both the set of predictors and the corresponding output for each subject, the MLP can learn out of these patterns of information. To do so, the MLP uses a learning algorithm while reducing an error function. On this matter, Haykin (2009) suggests that one common and simple method for the training of MLP is the back-propagation algorithm which happens in two stages. In the forward stage, the predictive weights of the network are calculated, and the input signal is transmitted through the layers until reaching the output. Then, in the second stage, an error signal is generated by comparing the output of the network with the expected value. This error signal is also propagated, layer by layer, but in the reverse direction. By doing so, the ANNs can adjust the value of the previous predictive weights, minimizing the error in each cycle, until a certain level of precision is reached.

During this back-propagation process, two parameters of the network are adjusted. The learning rate, which changes the value of the weights in each iteration of the learning process, and

the momentum, which increases the speed of the learning process by adding a fraction of the previous weight change to the current weight change (Atthoh-Okine, 1999). Furthermore, there are several activation functions that can be selected when implementing an MLP. The use of functions such as hyperbolic tangent or sigmoid allows the ANNs to identify complex and non-linear relationships between the inputs and the outputs (Somers & Casal, 2009).

2.2. Measures of the quality of MLPs in classification problems

There are two different ways to assess the quality of the classifications obtained after building MLPs models. The first one is to calculate an error function during the training phase; the interest is to compare the desired output with the real output provided by the model at each iteration of the training (Haykin, 2009; Garson, 2014). The second one is the use of well-known measures based on the confusion matrix of the classification, as in the case of the binary logistic regression analysis. In this respect, rates such as True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) should be used to calculate several measures of the quality of the MLPs (Musso et al., 2013). *Sensitivity* defined as TP/(TP+FN); *specificity* defined as TN/(TN+FP); *precision* defined as TP/(TP+FP); *accuracy* defined as (TP+TN)/(TP+FP+FN+TN); and *F1 score* defined as 2TP/(2TP+FP+FN).

2.3. Artificial neural networks to predict academic performance

The use of ANNs for predicting students' academic performance is still somewhat controversial in the literature. On the one hand, a first body of literature has focused on showing the obstacles that might arise for the use of ANNs as a prediction method of students' academic performance. The general lack of knowledge in the educational field regarding mathematical modeling, and machine-learning applications, has hindered the widespread use of these methods (Garson, 2014). In addition, several criticisms of the use of ANNs for the prediction of academic

performance can be found within this body of literature. The first one is the suggested impossibility to fully transfer the results of ANNs models into a set of rules because there is no precise information about the input-output relationship (Edelsbrunner & Schneider, 2013; Golino and Gomes, 2014). The second one is the absence of clear statistical parameters in the ANNs as they exist in traditional statistical techniques (Edelsbrunner & Schneider, 2013). Consequently, there is still a limited understanding of how the quality of the predictions resulting from models based on ANNs can be assessed. Finally, ANNs have been understood as "black boxes", where it is not possible to interpret the importance of the set of predictors when estimating an output variable (Benítez, Castro, & Requena, 1997).

On the other hand, a second body of literature addresses several misconceptions about the use of ANNs models and discussed possible resolutions of those issues. In relation to the unclear input-output relationship, Cascallar et al. (2015) explain that recent developments of ANNs have shown a more transparent use, which avoids the "black box" interpretation. In particular, the relationship between inputs and outputs is fully represented in the interactions which determine the predictive weights. Although this relationship is complex it exists in the model. Regarding the statistical parameters of ANNs, there are some good examples of the determination of confidence intervals and prediction intervals of the ANNs models (e.g., De Vieaux, Schumi, Schweinsberg, & Ungar, 1998; He & Li, 2011). Another way to determine the confidence intervals of the ANNs models is the use of least square estimations. Furthermore, it is also possible to use sensitivity analysis in order to determine the predictive weights of the inputs on the outputs in the ANNs models (Musso et al., 2013).

Notably, several limitations of the traditional statistical techniques can be solved by using ANNs. Firstly, educational researches face the challenge of analyzing large-scale data sets when

evaluating students' performance (Qianyin & Bo, 2015). Hence, there is a need for implementing methodological approaches, such as ANNs, for processing and explaining a vast amount of educational data. Secondly, traditional statistical techniques allow an acceptable level of prediction. Nevertheless, their capability for discovering undetected patterns that exist in the educational data is not as efficient as the capability of ANNs. As ANNs analyze more information represented by more cases and variables, better representation, and understanding of the complex interactions between predictors and predicted variable is achieved (Cascallar et al., 2015). Thirdly, while the use of only one predictive model could produce misleading predictions, ANNs rely on the use of several topologies and different learning paradigms, which enhances the quality of the prediction and produces more reliable interpretations (e.g., Dharmasaroja & Kingkaew, 2016).

2.4. Predictors of Academic Performance in Higher Education

A considerable number of studies have been conducted to identify which factors can explain students' academic performance in higher education. These factors have been classified in several theoretical categorizations. For example, McKenzie and Schweitzer (2001) propose four categories: academic, cognitive, demographic, and psychosocial factors. In their meta-analysis, Richardson, Abraham, and Bond (2012) propose six categories: demographics (i.e., sex, age, socio-economic status), traditional factors (i.e., prior academic achievement), personal traits, motivational factors, self-regulatory learning experiences, and psychosocial contextual influences. An additional classification including four categories, has been proposed by De Clercq, Galand, and Frenay (2017). They distinguish past performance, socio-economic status, self-efficacy beliefs, and study choice.

A comparison among these different classifications reveals that prior academic achievement and socio-economic status are two shared categories of predictors of students'

academic performance. However, two additional categories of predictors have also been explored in the literature: high school characteristics (e.g., Black, Lincove, Cullinane, & Veron, 2015; Win & Miller, 2005) and working status (e.g., Triventi, 2014; Yanbarisova, 2015).

2.4.1. Prior academic achievement

Prior academic achievement has been widely reported as one of the most important predictors of future academic performance at university (Li, Chen, & Duanmu, 2010; McKenzie & Schweitzer, 2001). More precisely, high school grade point average (HSGPA) and standardized entrance exams have been defined as predictors of academic performance in university (Pike & Saupe, 2002). The general trend in the literature has been to compare HSGPA with standardized entrance exams in their predictive validity of grade point average in university (GPA), finding that HSGPA is a stronger and valid predictor of academic performance in university (Chissom & Lanier, 1975; Cliffordson, 2008; Geiser & Santelices, 2007).

However, a more interesting conclusion has also been reported in these predictive studies. Although HSGPA and standardized entrance exams are related to GPA, they seem to predict either only cognitive (standardized entrance exams) or both cognitive and non-cognitive (HSGPA) elements of academic performance (Noble & Sawyer, 2002; Westrick, Le, Robbins, Radunzel, & Schmidt, 2015). Moreover, a more accurate prediction of GPA in university has been found when combining HSGPA and standardized entrance exams (e.g., Bridgeman, Pollack, & Burton, 2008). Therefore, the selection of both HSGPA and standardized entrance exams as predictors of academic performance in university has been highly recommended (Alnahdi, 2015; Mattern, Shaw, & Kobrin, 2011).

2.4.2. Socio-economic status (SES)

Socio-economic status (SES) has been extensively considered as a predictor of students' academic performance. Despite the lack of agreement on its definition, it can be argued that SES is usually conceptualized through the dimensions proposed by Mueller and Parcel (1981): economic, power, and prestige. Although one might consider that the predictive role of SES is captured by prior academic achievement, recent educational research has challenged this assumption. In this regard, Marks (2017) has suggested that their socio-economic conditions do not entirely explain the effect of prior academic achievement on students' academic performance in university at earlier stages of life. Consequently, it is still necessary to investigate the predictive role of both SES and prior academic achievement on academic performance in higher education.

2.4.3. High school characteristics

Previous research supports the idea that high school characteristics interact with students' socio-economic conditions to determine their academic performance in university. For instance, Wolniak and Engberg (2010) reported that students coming from higher family income groups also attended high schools with better infrastructure, and then had a higher academic performance in college than those students coming from low family income environments. In addition, students from low family income groups were also exposed to school violence, which was associated with a decrease in their academic performance in university. Similarly, results from Black et al. (2015) showed that students who came from high schools with higher average SES, more experienced teachers, district economies of scale, and fewer students with financial need, had a higher GPA in university. Therefore, it could be the case that high school characteristics interact with students' socio-economic conditions to reinforce the advantages of wealthier students (Wolniak and Engberg, 2010).

2.4.4. Working status

Triventi (2014) has proposed four approaches to explain the influence of students' working status on their academic performance. In the first one, the so-called zero-sum approach, differences in academic performance between working and non-working students can be expected after controlling for contextual variables. On the other hand, in the reconciliation approach, no major differences in academic performance should be expected, even after controlling for socio-economic characteristics. The negative selection-to-work approach suggests that differences in academic performance exist, but these would disappear after controlling for variables such as social background, ability, and motivation. Finally, the positive selection-to-work approach suggests that working students have more motivation than non-working students, which might compensate for the negative effect of working status on academic performance.

2.5. Standardized Tests as Indicators of Academic Performance

Standardized tests can be used in different ways including accountability, appraisal, and assessment (Morris, 2011). In particular, standardized testing assesses the level of knowledge in subject areas (e.g., mathematics, physics, chemistry) or performance in competencies (e.g., verbal, quantitative reasoning, writing) (Kuncel & Hezlett, 2007). Zucker (2003) distinguishes two types of standardized testing in relation to the way of reporting and interpreting their outcomes. Norm-referenced testing (NRT) aims to compare the results of a given test-taker to the rest of the test takers (reference group). To make this comparison possible, results are reported as percentile ranks or stanines. Criterion-referenced testing (CRT) seeks to report the level of accomplishment in terms of several performance level standards. Consequently, the level of performance (e.g., high, middle or low) is reported for each test taker.

Tests such as the SAT administered in the USA (Sackett et al., 2012) or the SweSAT administered in Sweden (Cliffordson, 2008) have been reported as predictors of future performance in university studies. As a unique case in higher education, standardized tests are used in Colombia not only to verify students' academic preparation before university but also to measure both generic and specific competencies acquired during their university studies. In this regard, the Colombian Institute for Educational Evaluation (*Instituto Colombiano para la Evaluación de la Educación*, ICFES) designs and manages two tests which are of importance for higher education: the SABER 11 test and the SABER PRO test.

The SABER 11 is a standardized test required for admission to higher education institutions in Colombia. The previous version of the SABER 11 test is a cluster of exams that assess students' level of achievement in eight subject areas at the end of high school: biology, physics, chemistry, mathematics, Spanish, English, social sciences, and philosophy. Each subject test consists of multiple-choice questions with only one correct answer and three distractors. Test results are reported using a scale from 0 to 100 for each subject resulting from an item response theory (IRT) analysis of the data. The new version of the SABER 11 test (introduced in 2014) assesses students' performance in the following sections: critical reading, mathematics, social sciences and civic competencies, natural sciences, and English, maintaining the same type of question and scoring from the previous version of the test (ICFES, 2019a).

The SABER PRO is a standardized test required for graduating from higher education institutions in Colombia. The exam assesses students' performance in the following generic competencies: critical reading, quantitative reasoning, written communication, civic competences, and English. Each section of the test consists of multiple-choice questions with only one correct answer and three distractors. In addition, the SABER PRO evaluates specific competencies such

as scientific thinking for science and engineering students; teaching, training and evaluating for education students; conflict management and judicial communication for law students; agricultural and livestock production for agricultural sciences students; medical diagnosis and treatment for medicine students; and organizational and financial management for business administration students. Score in each generic competency is reported through a scale going from 0 to 300 resulting from an IRT analysis of the data (ICFES, 2019b).

3. The Present Study

Predictive systems of academic performance based on ANNs have been used successfully in recent educational research in different educational settings (e.g., Cascallar & Musso, 2008; Musso & Cascallar, 2009; Turhan, Kurt, & Engin, 2013; Yıldız Aybek & Okur 2018). Nevertheless, three limitations have prevented the implementation of ANNs as predictive systems in higher education. In the first place, a systematic and generalizable way to *design* ANNs is still missing. In the second place, there is a limited understanding of how to *evaluate* the quality of the predictions derived from ANNs. In the third place, ANNs have been erroneously conceived as "black boxes", which makes it difficult to *interpret* the relationship between their inputs and outputs. To further extend the use of ANNs in educational research, the first objective of this study is to propose a methodological framework in order to address the limitations mentioned before. Such a methodological framework consists of three stages, namely, *design*, *evaluation*, and *interpretation*.

The second objective of this study is to give an illustration of the proposed multistage approach starting from the outcomes in standardized tests in Colombia. Therefore, MLPs are built in order to classify the outcomes in the quantitative reasoning section of the SABER PRO test in a cohort of 162,030 Colombian students. The selected predictors are prior academic achievement,

socio-economic status, high school characteristics, and working status. The purpose is to identify the predictive weight of each of these variables in the classification of the students' test outcomes as "high-performance" or "low-performance".

Considering all the above-mentioned reasons, the present study aims to answer the following research questions:

1. What can be learned from the systematic design of an MLP to predict students' academic performance?
2. How could we assess the quality of an MLP when used to classify students' academic performance?
3. What is the importance of prior academic achievement, socio-economic status, high school characteristics, and working status when used to classify students' academic performance?

4. Methodology

4.1. Description of the Participants

The Colombian Institute for Educational Evaluation (ICFES) provided the data for this study. The total sample included 162,030 students of both genders, females (60%) and males (40%), with a mean age of 23.5 years (*SD* =2.89 years), from private (64.4%) and public universities (35.6%) in Colombia. The students graduated from academic secondary schools (61.6%), both academic and technical secondary schools (18.1%), technical secondary schools (16.7%), and teacher-training secondary schools (3.6%). Furthermore, the students in the sample were studying one of several different academic programs: agronomy, veterinary, and related degrees (1.3%), arts (3.5%), economics, management, and accounting (25.4%), education (8.7%),

engineering, architecture, urban planning and related degrees (25.6%), health (11.4%), mathematics and natural sciences (2.1%), and social and human sciences (21%).

4.2. Measures

The predictors analyzed in this study were sorted based on their similarities, which led to defining the measures described in this section. Descriptive statistics for every single predictor are presented in Annex 1.

Prior academic achievement. Students' scores in seven subject areas (biology, physics, chemistry, mathematics, Spanish, social sciences, and philosophy) of the SABER 11 test (between the years 2006 and 2011) are selected as the operationalization of prior academic achievement.

Tuition fees. How students pay for their tuition fees describes this measure. Students pay their tuition fees through either one or several funding sources: parents, educational loan, own resources, and scholarship.

Students' socio-economic status. Information regarding parents' educational level, parents' occupation, and monthly family income are selected as indicators of students' socio-economic status (Sirin, 2005; Van Ewijk & Sleegers, 2010).

Students' home characteristics. The number of rooms, whether a student has a computer at home, internet access at home, type of accommodation (permanent or temporal), and home stratum are combined under this measure**.** Home stratum refers to the seven categories used by the Colombian government to classify households based on their physical characteristics and surroundings. The main reason behind this classification is to establish the price of public services hierarchically in each area (The World Bank, 2012).

Students' household status. This measure indicates whether a student is or not head of household and the number of persons the student has in his or her care.

Students' background information. Students' age when taking the SABER PRO test and students' gender define their background information.

High school characteristics. Several high school characteristics are part of this measure. These characteristics are high school academic calendar (indicated by the starting month of the academic year), high school "gender" (boys-only, girls-only, mixed), high school type (public or private), high school schedule (morning-only, afternoon-only, evening-only, weekend-only, ordinary), and high school diploma (academic, technical, academic-technical or teacher-training).

Working status. The number of weekly working hours is the indicator of students' working status (Beauchamp, Martineau, & Gagnon, 2016; Triventi, 2014).

University background. The number of academic semesters completed by the student when taking the SABER PRO test, university type (public or private), and students' academic program, are analyzed under this measure of university background.

Academic performance in higher education. Students' outcomes in the quantitative section of the SABER PRO test from the 2016 administration are chosen as the operationalization of academic performance in higher education. This choice of outcome data is due entirely to the information provided by the ICFES that could be successfully merged with the measures of prior academic achievement, SES, high school characteristics, and working status described earlier.

4.3. Analyses

This section begins by explaining which steps were followed to prepare the data for analysis. Subsequently, the multistage approach is described in detail, discussing each one of the proposed stages: *design, evaluation,* and *interpretation.*

4.3.1. Preparation of the data

Five steps were followed to prepare the data analyzed in this study. In step 1, students' outcomes were categorized as "high-performance" or "low-performance," according to the levels of performance of the SABER PRO test proposed by the ICFES and presented in Table 4. Therefore, the categorization of students' performance was based on performance standards instead of percentiles of the outcomes. The ICFES classifies students' scores from *level 1* to *level 4*, also providing general and specific descriptors of what students can do at each level. As the level increases the complexity of the descriptors also increases. A higher level encompasses the descriptors of the previous levels in addition to the descriptors of that particular level. *Level 1* describes the lowest students' performance, while *level 4* describes the highest students' performance. In this study, "high-performance" students were students from level 4 and "low-performance" students were students from level 1. In step 2, students were dummy coded as *belonging* or *not belonging* to each performance group. In step 3, the independent continuous variables were standardized (converted to z-score), so that their different scales could be comparable. In step 4, all the nominal and categorical variables were dummy coded. Each continuous variable, as well as each category from nominal and categorical variables, represents an input node for the MLPs. Consequently, there were 122 input nodes (input layer) and two output nodes (output layer) in each one of the MLPs. In step 5, the total sample was divided into training sample (80%, n=129,624) and testing sample (20%, n=32,406). Such a division has been previously suggested in the machine-learning literature (e.g., Suthaharan, 2016).

4.3.2. Multistage approach for implementing MLPs

NeuroSolutions 7.1.1 was the software tool chosen for conducting the analyses of this study. NeuroSolutions is a neural network package which combines a modular, icon-based design

interface with an implementation of advanced artificial intelligence and learning algorithms in a friendly to use interface. The multistage approach proposed in this study consisted of three stages, namely, *design, evaluation,* and *interpretation.* Each of these stages is described as follows.

The *design stage (stage 1)* aimed at systematically train and test several MLPs to classify students' academic performance. The parameters of the "best" model which maximize the correct classification in each performance group were determined after following the steps presented in Figure 1 and changing the parameters listed in Table 1. The reasoning behind these steps was to avoid a trial and error process. Moreover, it was intended to propose a set of steps which can be generalizable to different data sets. This systematic process led to the training and testing of 135 MLPs in each performance group, resulting in a total of 270 ANNs models. The error function, as well as different quality measures, were calculated for each model.

The *evaluation stage (stage 2)* assessed the classification resulting from the "best" model by exploring the relationship between its error and quality measures. The "best" model for classifying each performance group (identified in stage 1) was trained and tested 30 times again. Even though more iterations would imply more processing time, there was enough information after 30 iterations to carry out two specific analyses. In the graphical analysis, the error training graph was generated for the "best" model of each performance group. The average of the error (average of the mean squared error) and its standard deviation allowed the calculation of the confidence intervals for these graphs. In the correlational analysis, the relationship between quality measures and the root mean squared error (RMSE) was established using the information stored after the 30 iterations. It is important to clarify that the weights of the "best" models were randomly initialized at every new iteration. Therefore, the expected subsequent improvement of a previous model was avoided by generating a new model with no prior learning.

The *interpretation stage (stage 3)* sought to determine the contribution of the selected predictors to the classification of students' academic performance in each performance group. The predictive contribution of each variable for the classification of students' academic performance was analyzed as it is described next. First, the relative contribution of each variable was normalized, with 1 being the value of the most important variable, and the other variables having an importance value determined with respect to the most important one. Subsequently, the variables were grouped in the categories defined in section 4.2. The importance of each category was calculated by adding the predictive contribution of all the variables within that category. Finally, a comparison across the categories in each performance group was carried out.

5. Results

The results of this study are presented in accordance with the research questions. As the objectives of this study aim to provide a methodological framework to use ANNs to predict academic performance, results are shown for each one of the proposed stages.

5.1. Stage 1: Designing an MLP in a systematic way

The *design stage* aimed at systematically train and test several MLPs to classify students' academic performance. Annexes 2 and 3 present the value of the error and the quality measures of each of the 270 systematically trained and tested MLPs. The different effects found during the training and testing phases are described below.

5.1.1. Effects during the training phase

Four effects were identified during the systematic training of the MLPs. The first effect was related to the time of training. The time of training was the least when *softmax* was selected as the transfer function of the output layer regardless of the values of learning rate and momentum. This fact could be indicating that the *softmax* function converges faster towards the minimum error

than the other two output functions (sigmoid and linear sigmoid). However, less time in training does not guarantee lower values in the error function.

The second effect was related to the "shape" of the training. A "symmetric" training was found when *softmax* was selected as the output function. A symmetric training means that the same percentage of correctly classified cases was assigned by the MLP to the categories *belonging* and *not belonging* to each performance group. This fact is explained because *softmax* allows a categorical probability distribution in the classification, so it is possible to truly determine whether a case belongs or not to a category. In contrast, for the additional possible output functions, there was an "asymmetrical" classification. In other words, there was always a category with a larger percentage of correctly classified cases. This finding is in line with the idea of maximizing the correct classification of cases belonging to each performance group.

The third effect was the presence of a peak in the error function. Such a peak was only identified when selecting *softmax* as output function as well as low values of momentum (0.1, 0.2, 0.3) and learning rate (0.1). This peak in the error function might produce an unbalanced classification in the testing phase of the MLPs so that most of the cases would be incorrectly assigned to only one category. Figure 2 gives an illustration of the peak in the error function described earlier.

The fourth effect was the relationship between learning rate, momentum, and error. When either *sigmoid* or *linear sigmoid* were selected as output functions, the amount of error increased when the learning rate and momentum also increased. Such an increase in the error was larger for the linear sigmoid function than for the sigmoid function. However, the amount of error was less sensitive to changes in learning rates when the sigmoid function was chosen. In fact, similar levels of error were found for both performance groups when the momentum changed between 0.1 and

0.5. Conversely, levels of error changed more rapidly when momentum was larger than 0.6 ("high-performance" group) and 0.7 ("low-performance" group). Altogether, these findings might suggest that a *sigmoid* function converges more effectively towards the minimum than the other two output functions.

5.1.2. Effects during the testing phase

Results of the training phase (Annexes 2 and 3) revealed that low values of error during the training phase do not necessarily correspond to better quality measures in the testing phase. An explanation of this finding could be an unbalanced classification due to overtraining. In this context, overtraining can be understood as the lack of discrimination in the predictive system as a result of the inclusion of random pattern effects during the classification. Therefore, only one category (either "belonging" or "not-belonging") would have an extremely high value of correctly classified cases.

At the testing phase, the "best" model for the "high-performance" group was obtained when the following parameters were selected: hyperbolic tangent (activation function), sigmoid (output function), a learning rate of 0.1, and momentum of 0.7. In this case, the model correctly classified 82% of the cases. Similarly, the "best" model for the "low-performance" group was reached through the following parameters: hyperbolic tangent (activation function), sigmoid (output function), a learning rate of 0.2, and momentum of 0.2. In this case, the model correctly classified 75% of the cases.

5.2. Stage 2: Evaluating the quality of the MLPs

The *evaluation stage* assessed the classification resulting from the "best" model by exploring the relationship between its error and quality measures. Figure 3a shows the error function obtained for both performance groups following the procedure described in section 4.3.2.

According to Figure 3a, the error function converged faster to a minimum value in the "high-performance" group than in the "low-performance" group.

Figure 3b depicts the relationship between the error and the quality measures for both performance groups. It is important to note that two different axes are used simultaneously in Figure 3b, given the different scales of the quality measures (left axis) and RMSE (right axis). An analysis of how error and the quality measures were related (Figure 3b) showed interesting patterns among them. In the "high-performance" group, there was a positive and significant relationship between RMSE and sensitivity (r= .962, p<0.05); and a negative and significant relationship between RMSE and specificity (r= -.991, p<0.05), RMSE and accuracy (r= -.99, p<0.05), and RMSE and F1 score (r= -.99, p<0.05). Similarly, in the "low-performance" group, there was a positive and significant relationship between RMSE and sensitivity (r= .925, p<0.05); and a negative and significant relationship between RMSE and specificity (r=-.966, p<0.05), RMSE and accuracy (r= -.968, p<0.05), and RMSE and F1 score (r= -.898, p<0.05). As shown in Figures 3a and 3b, there is consistency in the classification provided by the MLPs. In other words, results in the classification can be expected to be similar over several iterations of testing for each of the designed MLPs.

5.3. Stage 3: Interpreting the predictive importance of the predictors

The *interpretation stage* sought to determine the contribution of the selected predictors to the classification of students' academic performance in both performance groups. Table 2 shows the normalized predictive importance of the predictors in the "high-performance" group. In this case, scores in six sections of the SABER 11 test (mathematics, social sciences, chemistry, biology, Spanish, and physics) and monthly family income were the indicators that contributed the most to the classification in the "high-performance" group. In addition, the predictors were grouped using

the categories presented in section 4.2. Figure 4 gives information on the predictive contribution of each category. Results indicated that prior academic achievement, students' SES, university background, and high school characteristics were the categories with the largest contribution to the students' classification in the "high-performance" group.

Table 3 shows the normalized predictive importance of the predictors in the "low-performance" group. In this case, scores in four sections of the SABER 11 test (mathematics, chemistry, biology, social sciences) and students' age were the indicators that contributed the most to the classification in the "low-performance" group. Similar to the "high-performance" group, the predictors were also grouped using the categories presented in section 4.2. Figure 5 gives information on the predictive contribution of each category. Findings suggested that prior academic achievement was the category that contributes the most to the students' classification in the "low-performance" group.

A comparison across Figures 4 and 5 reveal interesting results. Several variables provide significant proportions of information for the classification of the "high-performance" group. In contrast, predictors related to prior academic achievement, on their own, contribute the most to the classification of the "low-performance" group. SES indicators, as well as high school characteristics, had a larger predictive contribution in the "high-performance" group than in the "low-performance" group. Finally, working status did not have a considerable contribution to the classification in none of the analyzed performance groups.

6. Discussion

The objectives of this study were twofold: (1) to propose a multistage approach to use predictive systems based on ANNs in the prediction of academic performance in higher education; and (2) to give an illustration of the proposed multistage approach by classifying the outcomes of

the outcomes in the quantitative reasoning section of the SABER PRO test in a cohort of 162,030 Colombian students.

Regarding the first research question, the effects found during the systematic training and testing of the MLPs can be summarized as follows. To begin with, less time of training does not imply lower values of error in the classification. The convergence towards the minimum in the error function depends on the number of epochs and the transfer function in the output layer. In addition, when the interest is to maximize the number of correctly classified cases belonging to one specific category, a suitable choice for the transfer function in the output layer is the *sigmoid* function. The *sigmoid* function did not exhibit any peak during the training and converged more efficiently to the minimum of the error. Finally, it was found that due to the inclusion of random patterns (overtraining), the accuracy of the classification could decrease considerably.

Furthermore, it is not advisable to determine the values of ANNs parameters such as *learning rate* and *momentum* at random. A random selection is problematic because it would produce undesirable effects in the training of the ANNs such as an increase in the training time or reaching a local minimum in the error function. A random selection also ignores how additional parameters of the network, such as the transfer function of the output layer, influence how learning rate, momentum, and error are related. What seems to work best, instead, is to systematically vary the parameters of the ANNs under a controlled training and testing process, so the values that maximize the classification of interest can be appropriately determined. The steps proposed in this study for the design of the MLPs (Figure 1) can be easily replicated with any data set coming from different educational settings.

Regarding the second research question, results showed that it is possible to assess the quality of a classification obtained by a model based on ANNs. However, it is important to

highlight that lower levels of error during the training of an MLP do not necessarily imply better performance measures in the testing phase. The level of error during the training phase must be considered in order to verify the convergence of any given model. To assess the quality of a classification provided by an ANN, then, it is advisable to use different measures of quality and observe their stability over several iterations of testing. Measures of quality, such as the F1 score should always be calculated in addition to the traditional measures of quality. The F1 score allows a balance between the values of sensitivity and precision. Moreover, it can be argued that maximizing the F1 score value implies higher sensitivity and precision of the predictive system. Besides, the F1 score also makes it possible to consider various utility factors through the differential weighting of sensitivity and precision.

Regarding the third research question, findings of this study are of importance as they are revealing interesting remarks for the prediction of students' academic performance in higher education. Firstly, this study corroborates that prior academic achievement is the most important predictor of academic performance in university. While this finding is not completely surprising, what it is more interesting to highlight is the fact that prior academic achievement differently behaves when classifying students' academic performance as either high or low. More specifically, the findings of this study show that the predictive weight of prior academic achievement was 27% for the "high-performance" group, whereas it was 66.1% for the "low-performance" group. Such difference seems to indicate that lower levels of prior academic achievement would classify students as "low performers" regardless of the interaction with the other predictors.

Secondly, results from this study also indicate that students' SES contributes to explain their academic performance in university. This study also suggests the need for analyzing SES indicators separately by their level of measurement (i.e., individual, family or area level). Similar

to prior academic achievement, the predictive weight of students' SES is also different within each performance group. In particular, the predictive weight of students' SES was 26% for the "high-performance" group. In the case of the "low-performance" group, students' SES was found to have a predictive weight of 9.7%.

Thirdly, it was found that the contribution of high school characteristics to the classification of the "high-performance" group (12%) was considerably larger than the "low-performance" group (1.5%). Remarkably, this difference between performance groups was the largest within all the predictors analyzed in this study. One feasible explanation of this finding is that students pertaining to each performance group come from high schools with similar characteristics. Therefore, the predictive weight of high school characteristics could be somehow embedded in other predictors such as prior academic achievement or students' SES. In this respect, past educational research has concluded that high school characteristics interact with students' SES to reinforce the advantages of wealthier students (Wolniak and Engberg, 2010).

Fourthly, the predictive weight of working status was low for both performance groups. However, this result warrants special attention given that working status was merely assessed as the number of hours worked in a week. In this respect, past educational research (Wang, Kong, Shan, & Vong, 2010; Yanbarisova, 2015) has suggested that working status should not be only defined as a homogenous category (i.e., working or not) but also considering additional information on why students work and characteristics of their work while attending university.

Finally, this study supports the idea of incorporating information on students' academic program when analyzing their academic performance in university. Students' academic program provided valuable information for the students' classification in both performance groups. A similar assertion has been pointed out in the educational literature (e.g., De Clercq, Galand, and

Dupont, 2013; Hansen & Mastekaasa, 2006). One possible explanation of this fact is that different academic programs correspond to different assessment practices, irrespective of using the same grading scale. Therefore, it could be the case that the predictive power of the selected predictors of academic performance varies across the different disciplines of study.

6.1. Implications for Educational Research

The multistage approach proposed in this study provides a methodological framework for implementing predictive models based on ANNs. Because of the design, evaluation, and interpretation of several MLPs, this study addresses several limitations of the use of ANNs as the predictive system of academic performance. Furthermore, this study suggested a more transparent use of ANNs, which challenges the assertion that they are merely "black boxes". The use of a large data set provided more information to the ANNs, so a better representation and understanding of the complex interactions between predictors and predicted variables was achieved. The predictive importance of the selected predictors was analyzed according to their contribution to the classification of academic performance in each performance group. Moreover, it was possible to make comparisons among the categories of predictors between the performance groups. Therefore, it can be stated that meaningful and useful interpretations can be obtained from the ANNs models.

Besides, the use of ANNs represents a different paradigm for educational researchers when predicting educational outcomes. It can be argued, for instance, that the training process of ANNs gives the opportunity of building several different models that predict students' academic performance in a non-linear way. This is possible due to the systematic changes in the network parameters which allow the predictive system to learn from the data during each iteration. Unlike the traditional statistical techniques, where a hypothesis is tested in a data set with only one model, the use of a machine-learning approach makes it possible to consider multiple models iteratively.

Whereas the use of standardized testing is still controversial among educational researchers, this study supports the idea of their use as indicators of academic performance in higher education. Standardized tests as the SABER PRO test could be a key predictor of future academic performance in university studies at the postgraduate level.

6.2. Limitations

Although the present study contributes to the use of ANNs in educational research, several of its limitations must be acknowledged. First, prior academic achievement was found to contribute the most to the classification of students' academic performance. However, this predictor was only operationalized through the SABER 11 test outcomes because no information on students' grades from high school was available. Therefore, relevant information represented by a traditional and widely reported indicator such as high school grade point average (HSGPA) could not be included in this analysis. Secondly, all the information regarding students' socio-economic conditions was self-reported by the students when they sat the SABER PRO test. This fact could lead to an inaccurate measurement of such a complex construct as SES. Thirdly, the literature on machine learning coming from different disciplines suggests the use of a validation set to verify the performance of predictive models based on ANNs effectively. This suggestion implies the availability of two different data sets available for analysis. However, in this study, only one data set was provided, so it was not possible to generalize the results to a second data set.

6.3. Future research

Future research on the use of ANNs to predict academic performance could focus on both educational and methodological topics. Regarding educational topics, indicators of academic performance, such as retention or dropout rate, could be defined as outputs of predictive systems based on ANNs. In this way, teachers, institutions, and policymakers could allocate efforts to

improve the quality of higher education systems. Furthermore, the complex nature of the relationship between SES and academic performance could be better understood by using non-linear approaches such as ANNs. Finally, the use of predictive systems based on ANNs, could contribute to the development of educational data mining (EDM), which represents a growing and relevant research field within educational sciences.

Concerning methodological topics, there is still needed to define more precise and more generalizable heuristics for the use of ANNs. In this respect, there is still room to explore how several parameters of the networks such as the number of hidden layers or the number of neurons in each layer, might influence their performance. Furthermore, different topologies of ANNs could be used to perform tasks such as clustering. Thus, it is highly recommended to consider the use of additional topologies based on different learning paradigms, so that a better comprehension of the methodological aspects involved in the use of machine-learning techniques can be achieved. Finally, despite the good examples of the determination of confidence intervals and prediction intervals of the ANNs models (e.g., De Vieaux, Schumi, Schweinsberg, & Ungar, 1998; He & Li, 2011) such task still needs consideration and future research.

References

Alnahdi, G. H. (2015). Aptitude tests and successful college students: The predictive validity of the General Aptitude Test (GAT) in Saudi Arabia. *International Education Studies, 8,* 1-6.

Attoh-Okine, N. O. (1999). Analysis of learning rate and momentum term in backpropagation neural network algorithm trained to predict pavement performance. *Advances in Engineering Software, 30,* 291-302.

Ayán, M. N. R., & García, M. T. C. (2008). Prediction of university students' academic achievement by linear and logistic models. *The Spanish Journal of Psychology, 11,* 275-288.

Beauchamp, G., Martineau, M., & Gagnon, A. (2016). Examining the link between adult attachment style, employment and academic achievement in first semester higher education. *Social Psychology of Education, 19,* 367-384.

Benítez, J. M., Castro, J. L., & Requena, I. (1997). Are artificial neural networks black boxes? *IEEE Transactions on Neural Networks, 8,* 1156-1164.

Black, S. E., Lincove, J., Cullinane, J., & Veron, R. (2015). Can you leave high school behind? *Economics of Education Review, 46,* 52-63.

Bridgeman, B., Pollack, J., & Burton, N. (2008). Predicting grades in college courses: A comparison of multiple regression and percent succeeding approaches. *Journal of College Admission, 199,* 19-25.

Cascallar, E. C., & Musso, M. F. (2008). Classificatory stream analysis in the prediction of expected reading readiness: Understanding student performance. *International Journal of Psychology, 43,* 242-243.

Cascallar, E., Musso, M. F., Kyndt, E., & Dochy, F. (2015). Modelling for understanding and for prediction/classification the power of neural networks in research. *Frontline Learning Research, 2,* 67-81. doi: 10.14786/flr.v2i5.135

Chissom, B. S., & Lanier, D. (1975). Prediction of first quarter freshman GPA using SAT scores and high school grades. *Educational and Psychological Measurement, 35*, 461-463.

Cliffordson, C. (2008). Differential prediction of study success across academic programs in the Swedish context: The validity of grades and tests as selection instruments for higher education. *Educational Assessment, 13,* 56-75.

Daniel, B. K. (2017). *Big data and learning analytics in higher education*. Basel, Switzerland: Springer International Publishing.

De Clercq, M., Galand, B., Dupont, S., & Frenay, M. (2013). Achievement among first-year university students: An integrated and contextualized approach. *European Journal of Psychology of Education, 28*, 641-662.

De Clercq, M., Galand, B., & Frenay, M. (2017). Transition from high school to university: A person-centered approach to academic achievement. *European Journal of Psychology of Education*, 32, 39-59.

De Vieaux, R. D., Schumi, J., Schweinsberg, J., & Ungar, L. H. (1998). Prediction intervals for neural networks via nonlinear regression. *Technometrics, 40*, 273-282.

Dharmasaroja, P., & Kingkaew, N. (2016). Application of artificial neural networks for prediction of learning performances. In L. Maozhen, N.Xiong, Z. Tong, J.Du, C.Liu, K. Li, & L. Wang (Eds). *12th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD)* (pp.745-751). New Jersey, USA: IEEE.

Edelsbrunner, P., & Schneider, M. (2013). Modelling for prediction vs. modelling for understanding. *Frontline Learning Research, 1*, 99-101.

Fausett, L. V. (1994). *Fundamentals of neural networks*. New Jersey, USA: Prentice-Hall.

Field, A. P. (2009). *Discovering statistics using SPSS: (and sex and drugs and rock 'n' roll).* Los Angeles, California: SAGE Publications.

Garson, G. D. (2014). *Neural network models*. Asheboro, NC: Statistical Associates Publishers.

Geiser, S., & Santelices, M.V. (2007). *Validity of high-school grades in predicting student success beyond the freshman year: High-school record vs. standardized tests as indicators of four -year college outcomes* (Report CSHE.6.07). Retrieved from UC Berkeley: Center for Studies in Higher Education website: https://escholarship.org/uc/item/7306z0zf

Golino, H., & Gomes, C. (2014). Four machine learning methods to predict academic achievement of college students: A comparison study. *Revista E-PSI, 4*, 68-101.

Hansen, M. N., & Mastekaasa, A. (2006). Social origins and academic performance at university. *European Sociological Review*, *22*, 277-291.

Haykin, S. S. (2009). *Neural networks and learning machines* (Vol. 3). Upper Saddle River, NJ, USA: Pearson.

He, S., & Li, J. (2011). Confidence intervals for neural networks and applications to modeling engineering materials. In C. L. P. Hui (Ed.), *Artificial neural networks – Application* (pp. 337-360). Shanghai, China: InTech.

Huang, S., & Fang, N. (2010). Prediction of student academic performance in an engineering dynamics course: Development and validation of multivariate regression models. *International Journal of Engineering Education, 26*, 1008-1017.

ICFES (The Colombian Institute for Educational Evaluation) (2019a). Información de la prueba SABER 11. [Specifications about the SABER 11 test]. Retrieved from: http://www.icfes.gov.co/acerca-examen-saber-once

ICFES (The Colombian Institute for Educational Evaluation) (2019b). Información de la prueba SABER PRO. [Specifications about the SABER PRO test]. Retrieved from: http://www.icfes.gov.co/acerca-del-examen-saber-pro

Khalilzadeh, J., & Tasci, A. D. (2017). Large sample size, significance level, and the effect size: Solutions to perils of using big data for academic research. *Tourism Management, 62*, 89-96.

Kuncel, N. R., & Hezlett, S. A. (2007). Standardized tests predict graduate students' success. *Science, 315*, 1080-1081.

Kyndt, E., Musso, M, Cascallar, E. & Dochy, F. (2015). Predicting academic performance : The role of cognition, motivation and learning approaches. A neural network analysis. In V. Donche, S. De Maeyer, D. Gijbels, & H. van den Bergh (Eds). *Methodological challenges in research on student learning*, (pp. 55-76) Garant: Antwerp, Belgium

Li, G., Chen, W., & Duanmu, J. L. (2010). Determinants of international students' academic performance: A comparison between Chinese and other international students. *Journal of Studies in International Education, 14*, 389-405.

Marks, G. N. (2017). Is SES really that important for educational outcomes in Australia? A review and some recent evidence. *The Australian Educational Researcher*, *44*, 191-211.

Mattern, K. D., Shaw, E. J., & Kobrin, J. L. (2011). An alternative presentation of incremental validity: Discrepant SAT and HSGPA performance. *Educational and Psychological measurement, 71*, 638-662.

McKenzie, K., & Schweitzer, R. (2001). Who succeeds at university? Factors predicting academic performance in first year Australian university students. *Higher Education Research and Development*, *20*, 21-33.

Morris, A. (2011). Student standardized testing: Current practices in OECD countries and a literature review. *OECD Education Working Papers, 65*, 1-51.

Mueller, C. W., & Parcel, T. L. (1981). Measures of socioeconomic status: Alternatives and recommendations. *Child Development, 52*, 13-30.

Musso, M.F. & Cascallar, E.C. (2009). Predictive systems using artificial neural networks: An introduction to concepts and applications in education and social sciences. In M.C. Richaud & J.E. Moreno (Eds.). *Research in Behavioural Sciences:Vol 1* (pp. 433- 459). Argentina:CIIPME/CONICET.

Musso, M. F., Kyndt, E., Cascallar, E. C. & Dochy F. (2013). Predicting general academic performance and identifying the differential contribution of participating variables using artificial neural networks. *Frontline Learning Research, 1*, 42-71, doi:10.14786/flr.v1i1.13.

Nimon, K. F. (2012). Statistical assumptions of substantive analyses across the general linear model: a mini review. *Frontiers in Psychology, 3,* 1-5.

Noble, J., & Sawyer, R. (2002). Predicting different levels of academic success in college using high school GPA and ACT composite score. *ACT Research Report Series*, *143,* 1-26.

Osborne, J., & Waters, E. (2002). Four assumptions of multiple regression that researchers should always test. *Practical Assessment, Research & Evaluation, 8,* 1-5.

Pike, G. R., & Saupe, J. L. (2002). Does high school matter? An analysis of three methods of predicting first-year grades. *Research in Higher Education, 43,* 187-207.

Qianyin, X., & Bo, L. (2015). Multiple evaluation models for education based on artificial neural networks. *International Journal of Hybrid Information Technology, 8,* 1-10.

Richardson, M., Abraham, C., & Bond, R. (2012). Psychological correlates of university students' academic performance: A systematic review and meta-analysis. *Psychological Bulletin, 138,* 353-387.

Sackett, P. R., Kuncel, N. R., Beatty, A. S., Rigdon, J. L., Shen, W., & Kiger, T. B. (2012). The role of socioeconomic status in SAT-grade relationships and in college admissions decisions. *Psychological Science, 23,* 1000-1007.

Sirin, S. (2005). Socioeconomic status and academic achievement: A meta-analytic review of research. *Review of Educational Research, 75*, 417-453.

Somers, M. J., & Casal, J. C. (2009). Using artificial neural networks to model nonlinearity: The case of the job satisfaction job performance relationship. *Organizational Research Methods, 12,* 403-417.

Suthaharan S. (2016). Supervised Learning Algorithms. In: R. Sharda & S. Voß (Eds.), *Integrated Series in Information Systems: Vol.36. Machine learning models and algorithms for big data classification* (pp. 183-206). Boston, MA: Springer.

Tacq, J. (1997). *Multivariate analysis techniques in social science research. From problem to analysis.* London, UK: Sage.

The World Bank (2012). *Reviews of national policies for education: Tertiary education in Colombia 2012*. Paris: OECD Publishing.

Triventi, M. (2014). Does working during higher education affect students' academic progression? *Economics of Education Review, 41*, 1-13.

Turhan, K., Kurt, B., & Engin, Y. Z. (2013). Estimation of student success with artificial neural networks. *Education and Science, 38*, 112-120.

Van Ewijk, R., & Sleegers, P. (2010). The effect of peer socioeconomic status on student achievement: A meta-analysis. *Educational Research Review, 5*, 134-150.

Wang, H., Kong, M., Shan, W., & Vong, S. K. (2010). The effects of doing part-time jobs on college student academic performance and social life in a Chinese society. *Journal of Education and Work, 23*, 79-94.

Westrick, P. A., Le, H., Robbins, S. B., Radunzel, J. M., & Schmidt, F. L. (2015). College performance and retention: A meta-analysis of the predictive validities of ACT® scores, high school grades, and SES. *Educational Assessment, 20*, 23-45.

Win, R., & Miller, P. W. (2005). The effects of individual and school factors on university students' academic performance. *Australian Economic Review, 38*, 1-18.

Wolniak, G. C., & Engberg, M. E. (2010). Academic achievement in the first year of college: Evidence of the pervasive effects of the high school context. *Research in Higher Education, 51*, 451-467.

Yanbarisova, D. M. (2015). The effects of student employment on academic performance in Tatarstan higher education institutions. *Russian Education & Society, 57*, 459-482.

Yıldız Aybek, H. S. & Okur, M. R. (2018). Predicting achievement with artificial neural networks: The case of Anadolu university open education system. *International Journal of Assessment Tools in Education, 5*, 474-490.

Zucker, S. (2003). Fundamentals of standardized testing. Pearson Education Website. http://images.pearsonassessments.com/images/tmrs/tmrs_rg/Fundamentals_of_Standardized_Testing.pdf

Table 1. Fixed and changeable parameters of the MLPs

Fixed parameters	
Number of input nodes	122
Number of output nodes	2
Number of hidden layers	1
Activation function of the hidden layer	Hyperbolic tangent
Number of neurons in the hidden layer	50
Number of epochs for training	200
Type of learning	Online
Changeable parameters	
Output function (x 3)	Sigmoid, Linear sigmoid, and Softmax
Learning rate in the hidden layer (x 5)	Going from 0.1 to 0.5 (steps of 0.1)
Learning rate in the output layer	10% of the learning rate of the hidden layer *
Momentum (x 9)	Going from 0.1 to 0.9 (steps of 0.1)

* This value is set by default by NeuroSolutions.

Table 2. Predictive weights for high-performance group

Variable	Predictive weight	Normalized importance (%)
SABER 11 mathematics score	0.053	100%
SABER 11 social sciences score	0.043	81%
SABER 11 chemistry score	0.041	78%
SABER 11 biology score	0.041	78%
SABER 11 Spanish score	0.040	75%
SABER 11 physics score	0.032	59%
Monthly family income	0.022	42%
University type (Public national funding)	0.020	38%
Academic program (Mathematics and natural sciences)	0.020	37%
SABER 11 philosophy score	0.020	37%
Father's occupation (N.A.)	0.019	36%
Academic program (Engineering, architecture, urban planning, and related degrees)	0.018	33%
High school diploma (Technical)	0.016	30%
High school schedule (Ordinary)	0.016	30%
Academic program (Education)	0.015	28%
High school diploma (Academic)	0.015	27%
Father's education (Primary education incomplete)	0.014	26%
Academic program (Economics, management, and accounting)	0.014	26%
Mother's occupation (Unsalaried worker at family business)	0.013	25%
Age	0.013	24%
Stratum (6)	0.012	23%
Number of persons in charge	0.012	23%
University type (Private corporation funding)	0.012	23%
High school academic calendar (A)	0.012	22%
High school academic calendar (B)	0.012	22%
Scholarship (No)	0.011	21%
Scholarship (Yes)	0.011	21%
High school gender (Mixed)	0.011	21%
Academic program (Social and human sciences)	0.011	20%
Mother's education (Bachelor's degree completed)	0.011	20%
Mother's education (Bachelor's degree incomplete)	0.011	20%
Mother's occupation (N.A.)	0.010	19%
Father's education (Secondary education incomplete)	0.010	19%
Father's occupation (Employee at private sector)	0.010	19%
Stratum (1)	0.010	19%
Gender (Female)	0.010	18%
Father's occupation (Self-employed)	0.010	18%
Gender (Male)	0.010	18%
Mother's occupation (Retired)	0.009	18%
Mother's education (Secondary education completed)	0.009	17%
University type (Public municipalities funding)	0.009	16%

Stratum (2)	0.008	16%
Mother’s education (Technical or technological degree completed)	0.008	16%
High school gender (Girls-only)	0.008	16%
Permanent accommodation	0.008	15%
Academic program (Not specified)	0.008	15%
Own resources (Yes)	0.008	14%
Own resources (No)	0.007	14%
University type (Public departmental funding)	0.007	14%
High school schedule (Morning-only)	0.007	14%
Mother’s occupation (Employer)	0.007	13%
Education loan (No)	0.007	13%
University type (Private foundation funding)	0.007	13%
Father’s education (Unknown)	0.007	13%
Education loan (Yes)	0.007	12%
Father’s education (Secondary education completed)	0.006	12%
Stratum (3)	0.006	12%
Mother’s education (Primary education incomplete)	0.006	12%
Temporal accommodation	0.006	12%
Academic program (Health)	0.006	11%
Mother’s occupation (Domestic worker)	0.006	11%
Father’s occupation (Day laborer)	0.006	11%
High school type (Private)	0.005	10%
Academic semester	0.005	10%
High school type (Public)	0.005	9%
Working hours per week (Between 21 and 30)	0.005	9%
Mother’s occupation (Unemployed)	0.005	9%
Father’s education (Bachelor's degree incomplete)	0.005	9%
Working hours per week (Less than 10)	0.005	9%
Stratum (4)	0.005	9%
PC at home (Yes)	0.005	9%
Mother’s education (Postgraduate education)	0.004	8%
High school diploma (Teacher-training)	0.004	8%
PC at home (No)	0.004	8%
Father’s education (Technical or technological degree incomplete)	0.004	8%
Tuition fees from parents (Yes)	0.004	7%
Father’s occupation (Unemployed)	0.004	7%
Tuition fees from parents (No)	0.004	7%
Father’s occupation (Retired)	0.004	7%
Mother’s occupation (Self-employed)	0.004	7%
Mother’s occupation (Employee at private sector)	0.003	6%
Mother’s occupation (Public servant)	0.003	6%
Father’s education (Postgraduate education)	0.003	6%
Academic program (Arts)	0.003	6%
Mother’s education (Secondary education incomplete)	0.003	6%

Working hours per week (More than 30)	0.003	6%
Number of rooms in the house	0.003	6%
Mother's education (Primary education completed)	0.003	5%
Working hours per week (0)	0.003	5%
High school schedule (Afternoon-only)	0.003	5%
Mother's occupation (Unsalaried worker at non-family business)	0.002	5%
High school diploma (Academic and Technical)	0.002	4%
Internet at home (Yes)	0.002	4%
Working hours per week (Between 11 and 20)	0.002	4%
Father's education (Technical or technological degree completed)	0.002	4%
Father's occupation (Public servant)	0.002	4%
Father's occupation (Employer)	0.002	4%
Father's occupation (Domestic worker)	0.002	4%
Internet at home (No)	0.002	3%
Father's education (Primary education completed)	0.002	3%
Household of the family (No)	0.001	3%
Father's education (Bachelor's degree incomplete)	0.001	3%
Mother's occupation (Day laborer)	0.001	2%
Mother's education (No education)	0.001	2%
Academic program (Agronomy, veterinary and related degrees)	0.001	2%
High school schedule (Weekend-only)	0.001	2%
Household of the family (Yes)	0.001	2%
High school schedule (Evening-only)	0.001	1%
Father's occupation (Unsalaried worker at non-family business)	0.001	1%
High school academic calendar (F)	0.001	1%
Stratum (5)	0.001	1%
High school gender (Boys-only)	0.000	1%
Mother's education (Technical or technological degree incomplete)	0.000	1%
Father's occupation (Other activity or occupation)	0.000	1%
Stratum (0)	0.000	1%
Father's education (No education)	0.000	1%
Mother's occupation (Other activity or occupation)	0.000	1%
Father's occupation (Unsalaried worker at family business)	0.000	1%
High school diploma (Unknown)	0.000	0%
Mother's education (Unknown)	0.000	0%
University type (Special regime)	0.000	0%
High school schedule (Full day)	0.000	0%

Table 3. Predictive weights for low-performance group

Variable	Predictive weight	Normalized importance (%)
SABER 11 mathematics score	0.195	100%
SABER 11 chemistry score	0.143	73%
SABER 11 biology score	0.120	62%
SABER 11 social sciences score	0.106	55%
Age	0.105	54%
SABER 11 Spanish score	0.075	38%
Academic program (Engineering, architecture, urban planning and related degrees)	0.037	19%
SABER 11 physics score	0.021	11%
Mother's occupation (Unemployed)	0.020	10%
Academic program (Mathematics and natural sciences)	0.019	10%
Father's occupation (Domestic worker)	0.019	10%
Mother's education (Primary education completed)	0.015	8%
Academic program (Social and human sciences)	0.008	4%
Academic program (Education)	0.008	4%
Stratum (0)	0.007	4%
High school schedule (Ordinary)	0.004	2%
Academic program (Economics, management, and accounting)	0.004	2%
Mother's education (Technical or technological degree completed)	0.004	2%
High school schedule (Morning-only)	0.004	2%
Mother's education (Secondary education incomplete)	0.004	2%
Mother's occupation (Public servant)	0.003	2%
Father's education (Bachelor's degree incomplete)	0.003	2%
Mother's education (No education)	0.003	2%
Mother's education (Unknown)	0.003	2%
Father's education (Secondary education incomplete)	0.003	2%
Own resources (No)	0.003	2%
Father's education (Primary education incomplete)	0.003	2%
Working hours per week (Less than 10)	0.003	2%
Gender (Male)	0.003	2%
Working hours per week (Between 21 and 30)	0.003	1%
Father's occupation (Self-employed)	0.003	1%
Own resources (Yes)	0.003	1%
Mother's occupation (Employer)	0.002	1%
Stratum (4)	0.002	1%
University type (Public municipalities funding)	0.002	1%
Tuition fees from parents (No)	0.002	1%
Academic semester	0.002	1%
Tuition fees from parents (Yes)	0.002	1%
Gender (Female)	0.002	1%
Number of persons in charge	0.002	1%
High school schedule (Evening-only)	0.002	1%
High school diploma (Teacher-training)	0.002	1%

Father’s occupation (Day laborer)	0.002	1%
High school gender (Boys-only)	0.002	1%
Stratum (1)	0.001	1%
Father’s education (Unknown)	0.001	1%
University type (Private foundation funding)	0.001	1%
SABER 11 philosophy score	0.001	1%
Father’s occupation (Other activity or occupation)	0.001	1%
PC at home (No)	0.001	1%
Father’s occupation (Employer)	0.001	0%
Mother’s occupation (Employee at private sector)	0.001	0%
Father’s occupation (Retired)	0.001	0%
Working hours per week (More than 30)	0.001	0%
Father’s occupation (Employee at private sector)	0.001	0%
Stratum (6)	0.001	0%
Mother’s education (Bachelor’s degree incomplete)	0.001	0%
Stratum (2)	0.001	0%
Academic program (Agronomy, veterinary, and related degrees)	0.001	0%
High school schedule (Afternoon-only)	0.001	0%
Father’s occupation (Unsalaried worker at family business)	0.000	0%
Working hours (0)	0.000	0%
University type (Public departmental funding)	0.000	0%
High school type (Private)	0.000	0%
High school type (Public)	0.000	0%
PC at home (Yes)	0.000	0%
Monthly family income	0.000	0%
High school academic calendar (F)	0.000	0%
Mother’s occupation (Unsalaried worker at non-family business)	0.000	0%
Education loan (No)	0.000	0%
Mother’s occupation (Day laborer)	0.000	0%
Stratum (5)	0.000	0%
Education loan (Yes)	0.000	0%
High school gender (Mixed)	0.000	0%
Scholarship (No)	0.000	0%
Scholarship (Yes)	0.000	0%
High school diploma (Academic and Technical)	0.000	0%
Father’s education (Technical or technological degree incomplete)	0.000	0%
Academic program (Arts)	0.000	0%
High school diploma (Academic)	0.000	0%
Father’s education (No education)	0.000	0%
High school gender (Girls-only)	0.000	0%
Mother’s occupation (Self-employed)	0.000	0%
Mother’s education (Postgraduate education)	0.000	0%
Internet at home (Yes)	0.000	0%
Mother’s education (Bachelor’s degree completed)	0.000	0%

Household of the family (Yes)	0.000	0%
Mother's occupation (N.A.)	0.000	0%
University type (Public national funding)	0.000	0%
Father's occupation (N.A.)	0.000	0%
University type (Private corporation funding)	0.000	0%
Stratum (3)	0.000	0%
Internet at home (No)	0.000	0%
Household of the family (No)	0.000	0%
Father's education (Primary education completed)	0.000	0%
Working hours per week (Between 11 and 20)	0.000	0%
Mother's education (Primary education incomplete)	0.000	0%
High school diploma (Technical)	0.000	0%
Mother's occupation (Other activity or occupation)	0.000	0%
Mother's occupation (Domestic worker)	0.000	0%
Father's education (Postgraduate education)	0.000	0%
Father's education (Technical or technological degree completed)	0.000	0%
Mother's occupation (Retired)	0.000	0%
Mother's education (Technical or technological degree incomplete)	0.000	0%
Father's occupation (Unemployed)	0.000	0%
High school academic calendar (B)	0.000	0%
Temporal accommodation	0.000	0%
Father's occupation (Unsalaried worker at non-family business)	0.000	0%
Academic program (Not specified)	0.000	0%
High school schedule (Weekend-only)	0.000	0%
Father's occupation (Public servant)	0.000	0%
Father's education (Bachelor's degree completed)	0.000	0%
High school diploma (Unknown)	0.000	0%
Number of rooms in the house	0.000	0%
Academic program (Health)	0.000	0%
High school academic calendar (A)	0.000	0%
University type (Special regime)	0.000	0%
Mother's occupation (Unsalaried worker at family business)	0.000	0%
Mother's education (Secondary education completed)	0.000	0%
Permanent accommodation	0.000	0%
Father's education (Secondary education completed)	0.000	0%
High school schedule (Full day)	0.000	0%

Table 4. Performance levels for the quantitative reasoning section of the SABER PRO test

Performance levels	General descriptor	Specific descriptors
Level 1 (Score between 0 and 125)	A student at this level identifies explicit information from a single source associated with everyday contexts, which is presented in tables or bar charts containing little data or involve two variables at most.	A student at this level might: • Establish relationships of similarity and order based on the provided information. • Represent in different ways the information contained in tables and graphs.
Level 2 (Score between 126 and 155)	In addition to the above, a student at this level identifies and interprets explicit information from various sources, which is presented in tables and bar graphs, while using simple arithmetic procedures from the given information.	In addition to the previously described, a student at this level: • Identifies and extracts explicit information presented in tables and bar graphs. • Represents information contained in bar charts in other record types. • Formulates strategies, validates simple procedures and solves problems in everyday contexts related to the use of money, business operations, etc., which require the use of: - One or two operations, such as addition, subtraction or multiplication. - Distributive property of the product with respect to the sum.
Level 3 (Score between 156 and 200)	In addition to the above, a student at this level extracts implicit information contained in unusual representations associated with the same situation and from a single source of information, argue the validity of procedures, and solve problems using models that combine arithmetic, algebraic, variational and random.	In addition to the previously described, a student at this level: • Identifies and extracts relevant, explicit or implicit information presented in unusual graphs, such as stacked bar charts, circular diagrams, etc. • Identifies differences between data representations associated with the same context. • Forecasts results, indicating a single value or a possible interval, based on trends in the presented data. • Proposes strategies and solves problems using percentage calculations, conversion of standard units, simple averages, basic notions of probability or counts that use the principles of addition and multiplication, in few steps or calculations.
Level 4 (Score between 201 and 300)	In addition to the above, a student at this level identifies and uses implicit information contained in unusual representations from various sources of information to understand a problem situation; argues the validity of procedures and uses them to solve problems,	In addition to the previously described, a student at this level: • Recognizes the meaning of arithmetic expressions given in the context of solving a problem. • Establishes and uses reference points in the plane, making use of notions of parallelism and rotations.

	deciding which one is the most appropriate.	• Proposes representations based on the manipulation and transformation of relevant data in contexts with one or more sources of information. • Proposes strategies and solves problems, in contexts with implicit information, using conversion of non-standard units, operations with decimals and the concept of proportionality and rule of three. • Identifies and corrects errors in proposed procedures as a solution to a problem. • Solve problems that require multiple operations or approximations as part of the solution process. • Validate and compare solution procedures of the same problem and the obtained solutions.

Source: The Colombian Institute for Educational Evaluation (ICFES)

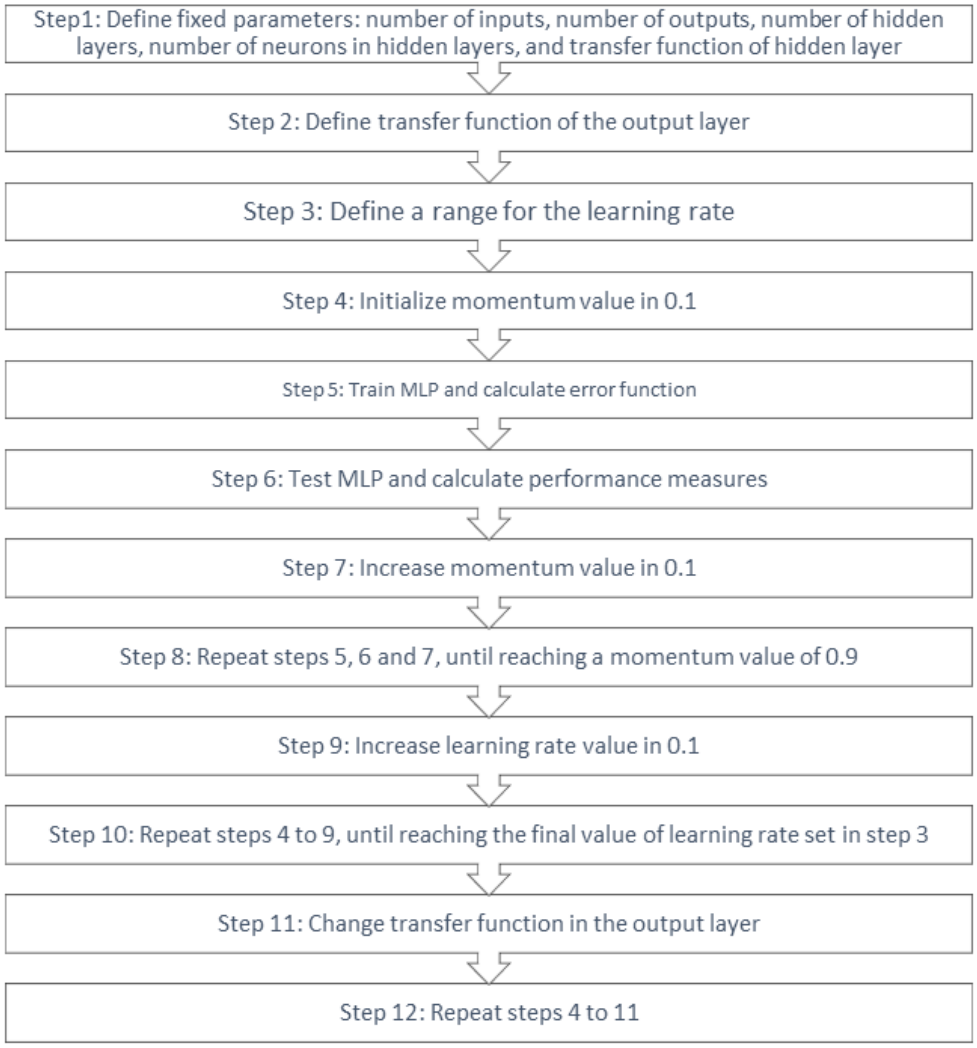


Fig.1 Steps for designing an MLP in a systematic way

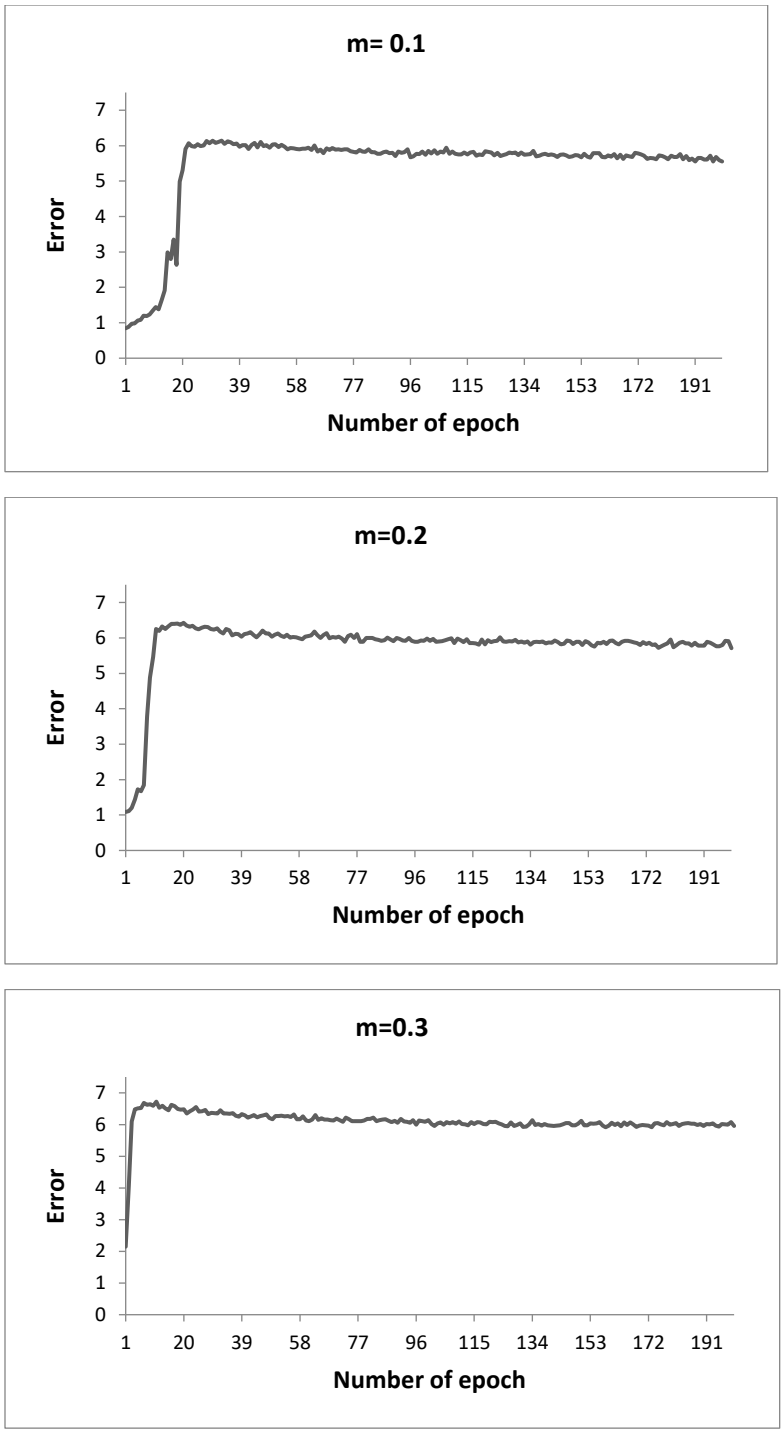


Fig.2 Peak in the error function during the training of the MLPs

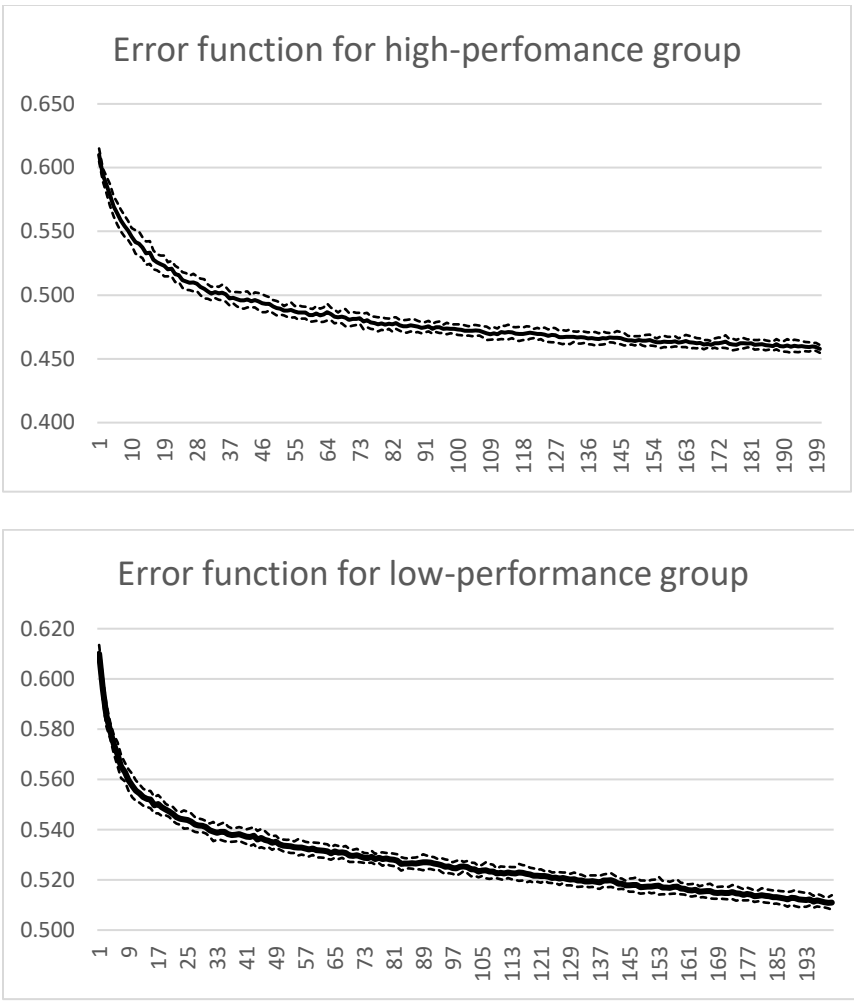
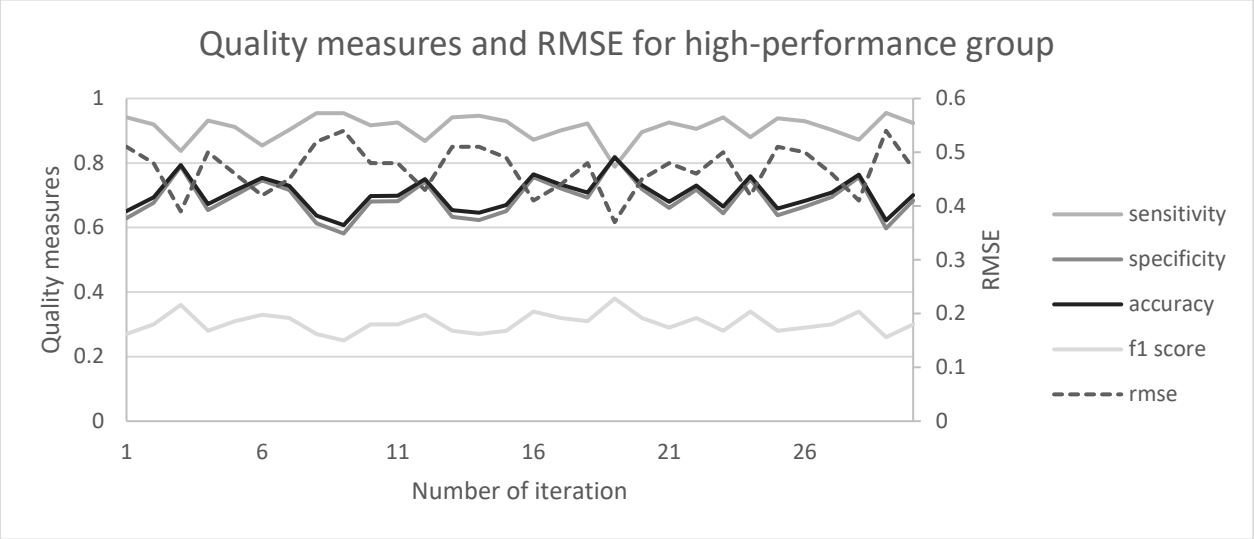


Fig.3a Error function for high-performance and low-performance group



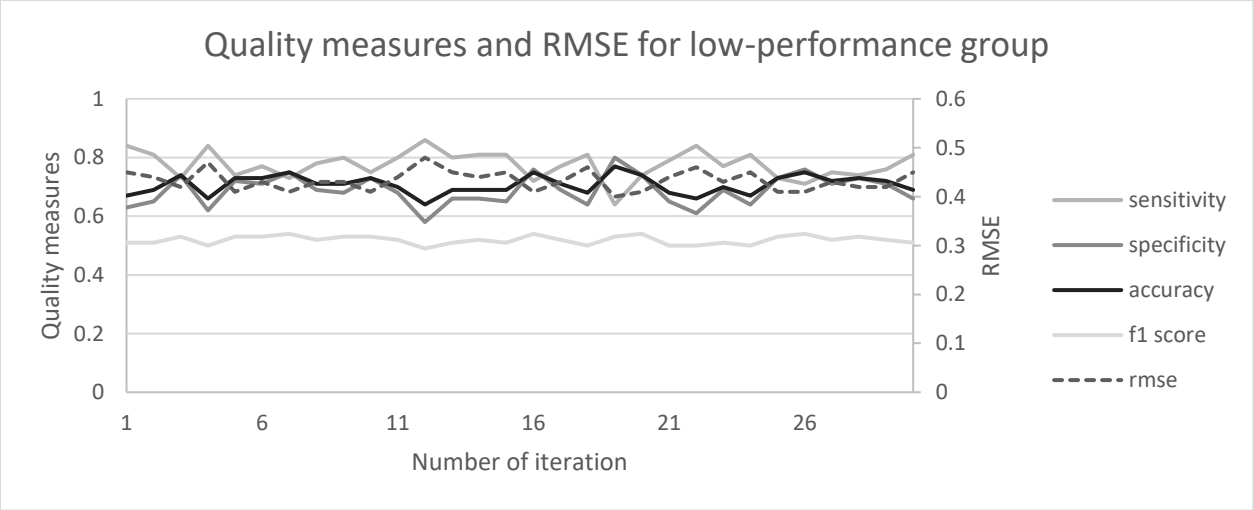


Fig. 3b RMSE and quality measures for both performance groups

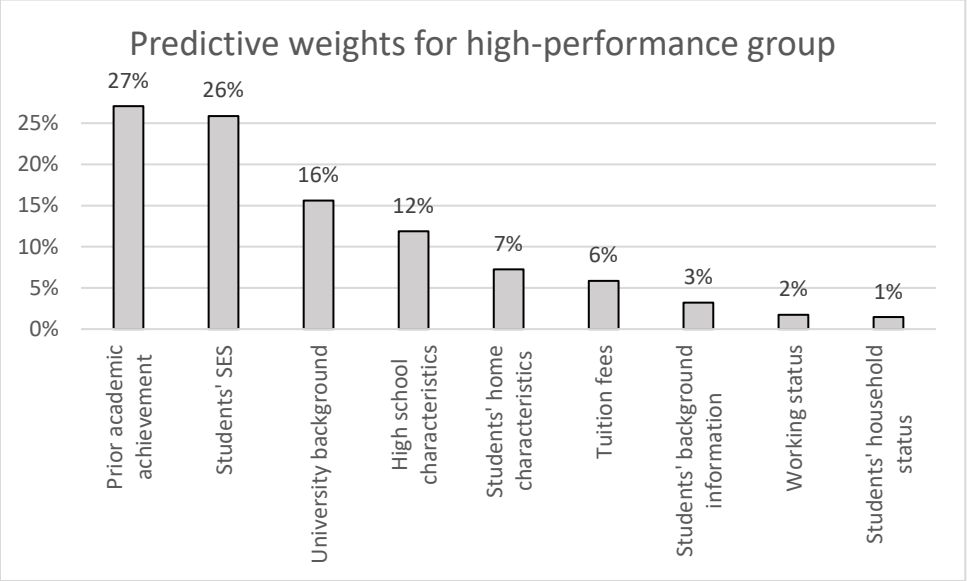


Fig.4 Categories of predictors for high-performance group

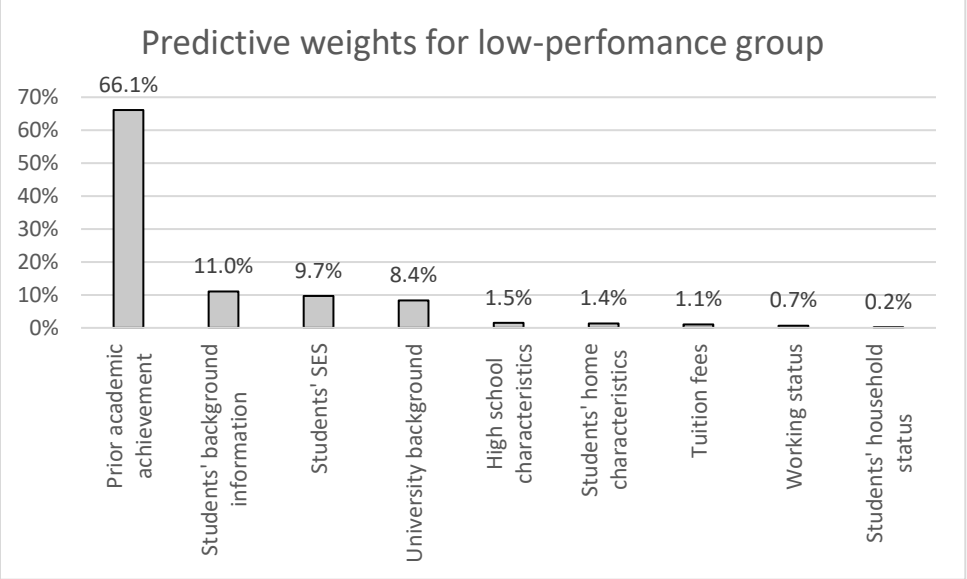


Fig.5 Categories of predictors for low-performance group

Chapter 4

The use of radial basis function neural networks (RBFNNs) to predict academic performance in higher education

Abstract

While educational researchers know regression analysis widely, its use still copes with several important limitations. In particular, the difficulty in interpreting regression coefficients derived from categorical predictors, as well as the undesired effects of violating statistical assumptions, highlight the need for exploring a different prediction approach. This methodological study shows the potential of the radial basis function neural network (RBFNNs) as a predictive tool for educational data. As an illustration case, RBFNNs are used to predict the academic performance in a cohort of 162,030 Colombian university students based on their prior academic achievement, socio-economic status (SES), high-school characteristics, and working status. Findings suggest that RBFNNs are easy to design and capable of satisfactorily represent and interpret the relationship between academic performance and its predictors, regardless of their level of measurement and statistical distribution. As such, RBFNNs provides a promising alternative to carry out predictions of educational data.

Keywords: neural networks, radial basis function, academic performance, higher education.

Introduction

The prediction of academic performance in higher education has shown a positive impact on teachers, students, and institutions (Musso, Kyndt, Cascallar & Dochy, 2013). For instance, it has contributed to evaluate the quality of teaching and learning (Lau, Sun, & Yang, 2019), and to identify which are the causes of dropping out as well as which students are at risk of dropping out (Berens, Schneider, Görtz, Oster, & Burghoff, 2018). Different predictive models have been proposed to analyze dichotomous indicators of academic performance, such as dropout, retention, or graduation (e.g., DeShields Jr., Kara, & Kaynak, 2005; Kara & DeShields Jr., 2004; Webber & Ehrenberg, 2010). Similarly, predictive models have been developed to estimate the weight of continuous indicators, for example, grade point average (GPA) (e.g., Cohn, Cohn, Balch, & Bradley Jr., 2004; Rothstein, 2004) or standardized test outcomes.

Concerning the methodological perspective, the general trend in the literature has been to use traditional statistical techniques such as multiple linear regression (MLR), to predict academic performance (e.g., Bonsaksen, 2016; De Clercq, Galand, Dupont, & Frenay, 2013; Gerken & Volkwein, 2000; Zheng, Saunders, Shelley, Mack, & Whalen, 2002). Four assumptions need to be fulfilled regarding the *residuals* of any MLR model: the absence of outliers, independence, normality, and homogeneity of variance (Chan, 2004; Osborne & Waters, 2002; Tacq, 1997; Williams, Gómez, & Kurkiewicz, 2013). Nevertheless, these assumptions are rarely verified in the studies aiming to predict academic performance. Consequently, results coming from this body of research are often methodologically biased.

Furthermore, although there are non-linear regression methods available, two limitations about their use should be acknowledged. First, the number of ways to combine the parameters in the regression equation is very large, which might convert the fitting of the non-linear model in a

trial and error task. Second, it is rather common to estimate the variance accounted by a non-linear model through the coefficient of determination (R^2). However, in a non-linear model, R^2 is not the correct choice given that the total variance of the model is not equal to the explained variance plus the error variance (Spiess & Neumeyer, 2010).

A powerful alternative for building predictive models and overcoming the limitations of traditional statistical techniques is the use of artificial neural networks (ANNs). In this respect, recent educational research (e.g., Ever & Dimililer, 2018; Musso, Kyndt, Cascallar & Dochy, 2013; Paik, 2000; Paliwal & Kumal, 2009; Teshnizi & Ayatollahi, 2015) has successfully built predictive models based on ANNs. One of the major contributions of this body of research has been to show that through ANNs it is possible to model non-linear relationships between variables, to process large volume of educational data, and to increase the accuracy of the prediction (Musso, Kyndt, Cascallar & Dochy, 2012, 2013; Kyndt, Musso, Cascallar & Dochy, 2015).

Among the different topologies of artificial neural networks, the radial basis function neural networks (RBFNNs) are a relatively new and useful type of artificial neural network used to predict a continuous variable (Haykin, 2009). First proposed by Broomhead and Lowe in 1988, RBFNNs have demonstrated advantages over the traditional artificial neural network models: less training time, a simpler structure of just three layers, and a high mapping capability (Jeatrakul & Wong, 2009). Given these advantages of RBFNNs, one might wonder whether their use could contribute to developing predictive systems of academic performance in higher education.

Considering all the mentioned above, this study is carried out to show the potential of RBFNNs as a predictive tool for educational data. The first objective of this study is to design RBFNNs and explore their possible architectures to best predict the academic performance of a cohort of 162,030 Colombian university students. The measurement of academic performance is

the scores of the quantitative reasoning section of the SABER PRO test, a standardized test that university students must sit in Colombia. The second objective of this study is to establish the predictive importance of prior academic achievement, socio-economic status, high-school characteristics, and working status when selected as predictors. Such a selection of well-established predictors in the educational literature (e.g., McKenzie, & Schweitzer, 2001; Shahiri Husain, & Rashid, 2015; Wang, Kong, Shan, & Vong, 2010) allows developing a more valid predictive model.

State of the art

This section presents a critical analysis of the use of regression models for the prediction of academic performance, as a starting point for the need for alternative methodological approaches. Next, artificial neural networks (ANNs) are defined. An emphasis is made on the radial basis function neural networks (RBFNNs), and their use for the prediction of academic performance in higher education.

Regression analysis for the prediction of academic performance

Regression analysis is a widely applied statistical technique for modeling the relationship between dependent and independent variables. The most common type of regression is linear regression, which models a linear relationship between one *dependent variable* and one (simple linear regression) or several (multiple linear regression) *independent variables* (Jeon, 2015). Through multiple linear regression (MLR), it is also possible to calculate the amount of variance of the dependent variable explained by the independent variables (Field, 2009; Jeon, 2015; Tacq, 1997). As in the case of other statistical techniques, the residuals of the MLR models (i.e., differences between the expected and the predicted values) need to fulfill assumptions regarding

the absence of outliers, independence, normality, and homogeneity of variance (Chan, 2004; Osborne & Waters, 2002; Tacq, 1997; Williams, Gómez, & Kurkiewicz, 2013).

A considerable number of studies have used MLR to analyze academic performance in higher education (e.g., Bonsaksen, 2016; De Clercq et al., 2013; Gerken, 2000; Zheng et al., 2002). The results derived from this body of research can be summarized as follows. Firstly, the percentage of explained variance of academic performance ranges between 25% and 48%. Secondly, academic performance in higher education is a multi-causal phenomenon that requires the consideration of several different predictors. Thirdly, students' characteristics such as cognitive factors, socio-economic status (SES), and psychosocial factors have been identified as predictors of their future academic performance.

Despite its enormous contribution to understanding academic performance in higher education, some shortcomings of the MLR approach warrant special attention. To begin with, several predictors of academic performance are measured using nominal or categorical scales. Dummy coding allows the inclusion of either nominal or categorical variables in predictive models based on MLR (Muijs, 2004). However, in such cases, the interpretation of the regression coefficients in MLR is not straightforward. This fact makes it difficult to entirely understand the predictive models based on MLR.

It is also the case that several statistical assumptions for the proper implementation of MLR models are not always met. The results of these violations are type I and type II errors and inadequate estimation of the effect sizes of statistical parameters (Nimon, 2012; Osborne & Waters, 2002). Consequently, the findings drawn from predictive models based on MLR might lead to inappropriate conclusions.

Furthermore, when predicting academic performance, it is rather common to select independent variables which are highly related among each other. In the MLR case, such linear relationships among predictors have been defined as multicollinearity. When there is multicollinearity, the sign of the regression coefficient of a given predictor might change from the corresponding sign of the correlation coefficient between that predictor and the dependent variable; moreover, regression coefficients could have larger standard errors (Alin, 2010). Therefore, multicollinearity could lead to less precise predictive models of academic performance.

To sum up, there are at least three major drawbacks of MLR when predicting academic performance, namely, difficulty in interpreting dummy coded variables, the unfulfillment of statistical assumptions, and multicollinearity. As MLR might not be the best approach for predicting academic performance, it is still necessary and desirable to explore a different methodological approach to achieve such a task in a valid and interpretable way.

Artificial neural networks: A short overview

Artificial neural networks (ANNs) are layered structures capable of representing complex interactions among a set of variables. Each layer contains processing units called neurons, which are related through specific weights indicating the strength of the connection (Fausset, 1994). There are at least three layers in an artificial neural network. The input layer receives the independent variables, the hidden layer maps the input onto the output, and the output layer returns the dependent variable (Musso et al., 2012, 2013; Kyndt et al., 2015).

ANNs can be used to solve several prediction tasks, such as classification or regression. Using the information provided by the set of predictors and their corresponding output, ANNs can learn out of these patterns of information (Marshall & English, 2000). To do so, the ANNs use a

learning algorithm while reducing an error function according to the information provided to the network (Haykin, 2009). In general, any artificial neural network could be characterized by (1) its arrangement of connections between the neurons (called its topology or architecture), (2) its approach to obtaining the strength or the weights on the connections (called its training or learning algorithm), and (3) its activation functions (Fausett, 1994).

ANNs have several advantages over traditional statistical models. A first advantage is that ANNs facilitate the analysis of vast volumes of information despite the level of measurement of the data (Garson, 2014; Musso et al., 2013). A second advantage is that ANNs do not require assumptions of normality, linearity, and completeness (Kent, 2009; Garson, 2014). A third advantage is that ANNs are robust predictive systems even if the selected predictors are highly related among them (Pao, 2008). Hence, the use of ANNs could also help to overcome some of the drawbacks of MLR when predicting academic performance discussed earlier (difficulty to interpret dummy coded variables, the unfulfillment of statistical assumptions, and multicollinearity).

Radial basis function neural networks (RBFNNs)

A radial basis function (RBF) is a mathematical function in which its value depends only on the distance from the origin or a different point known as a center (Orr, 1996). A typical and widely known example of an RBF is the Gaussian function, which is normally defined in terms of a *center* and a *width* (Dash, Behera, Dehuri & Cho, 2016; Garson, 2014; Haykin, 2009). Interestingly, when several Gaussian functions are first stretched and compressed, and then summed, the resulting sum can approximate any other function (Kriesel, 2005). To put it differently, a 'linear combination' of Gaussian functions can approximate any given function. Bearing this idea in mind, it is possible to design artificial neural networks to model any given data

distribution pattern. Such is the case of the radial basis function neural networks (RBFNNs) proposed by Broomhead and Lowe in 1988. RBFNNs are characterized by requiring less time for training, having a more straightforward structure with only one hidden layer, and showing a good mapping capability (Jeatrakul & Wong, 2009). Given the complex relationship existing between academic performance and its predictors, the former defining features of RBFNN could help to develop predictive systems of academic performance in higher education.

The architecture of the RBFNNs

Figure 1 shows the layered architecture of the RBFNNs. According to Haykin (2009), there are three layers in the RBFNNs:

1. *Input layer*, which connects the network to its inputs. The size of this layer is the number of inputs of the network.
2. *Hidden layer*, which applies radial basis functions to the information coming from the input layer. In the hidden layer, no weights are connecting the input layer and the units (i.e., radial basis functions). The number of units in the hidden layer could be equal to the number of exemplars in the training data set. However, in practice, the size of the hidden layer is smaller than the size of the training data set.
3. *Output layer*, which calculates the weights connecting the units from the hidden layer and the units in the output of the network. In the output units, a linear function takes the sum of such weights to provide the output of the network. Usually, there is only one unit in the output layer.

Training of the RBFNNs

One major feature of the RBFNNs is the existence of well-defined algorithms for their training (Dash et al., 2016; Garson 2014; Haykin, 2009). In general, the training process of RBFNNs occurs in two steps (Benoudjit & Verleysen, 2003; IBM, 2019; Schwenker, Kestler, & Palm, 2002). In the first step, the parameters of each basis function (center and width) are calculated using clustering methods. The specification of the number of units (i.e., number of radial basis functions) in the hidden layer of the RBFNNs is very critical. When the number of units is too low, the network will not show accurate results; conversely, when the number of units is too large, there might be overfitting. Thus, the network will not have good generalization capability (Huang & Fang, 2013). In the second step, the weights applied to the outputs of the given radial basis functions are determined through ordinary least-squares regression.

In brief, four parameters should be defined when training the RBFNNs: (1) the type of radial basis function (generally Gaussian function) in the hidden layer; (2) the number of units (i.e., number of radial basis functions) in the hidden layer; (3) the center and the width of the radial basis functions (calculated by the training algorithm); and (4) the weights of the output layer (calculated by the training algorithm).

Accuracy of the RBFNNs

The accuracy of the predictions of the RBFNNs models can be calculated as the difference between the expected value and the predicted value through the correlation coefficient or the mean squared error (Areerachakul & Sanguansintukul, 2009). One additional measure of the accuracy of the RBFNNs is the relative error defined in Equation 1, where y^m indicates the expected value, $\hat{y}^m$ is the corresponding predicted value and $\bar{y}$ is the mean of the dependent variable (IBM, 2019).

$$\frac{\sum_{m=1}^{M}(y^m - \hat{y}^m)^2}{\sum_{m=1}^{M}(y^m - \bar{y})^2} \quad (1)$$

RBFNNs for the prediction of academic performance in higher education

While a few studies have used RBFNNs to predict academic performance, they do show promising results. To begin with, Kayri and Çokluk (2016) build a predictive model based on RBFNNs to predict procrastination among 1271 university students in Turkey. A statistically significant correlation of 0.896 was found between the observed and the predicted data in the RBFNN model. A "general procrastination tendency" was found to be the most important predictor in the RBFFN model. Also, Haviluddin, Sunarto, and Yuniarti (2014) compared predictive models based on linear regression and RBFNN, which included learning motivation as a predictor of students' GPA. A smaller value for the mean squared error (MSE) was found in the RBFNN model. Therefore, RBFNN was reported as a more accurate technique to predict students' GPA. Finally, Huang and Fang (2013) predicted the final exam score in an engineering dynamics course through different mathematical models, including RBFNNs. The prediction was based upon the combination of several predictors such as students' GPA, grades obtained in four pre-requisite courses, and scores on the mid-term exams. Results demonstrated that predictive models based on RBFNNs had a percentage of accurate predictions (i.e., the weighted difference between the predicted final exam score and the observed final exam score) between 51% and 64%.

An illustrative case: predicting academic performance in higher education

Higher education aims at acquiring both generic and specific competencies (Sadler, 2013). Therefore, there is an increasing interest among academics, employers, and policymakers for helping students to develop competencies such as critical thinking, analytical reasoning, or

problem-solving (Shavelson, 2010). The acquisition of competencies is also one of the most common indicators of academic performance (York, Gibson, & Rankin, 2015). Although the assessment of competencies in higher education is rather complicated, the use of standardized tests such as the Graduate Medical School Admissions Test (GAMSAT) in Australia, the SABER PRO test in Colombia, or the College English Test (CET) in China has made it possible to achieve that task. In particular, the SABER PRO test is the standardized test used in Colombia to assess the level of competencies acquired by university students. More specifically, SABER PRO assesses students' performance in the following areas: critical reading, quantitative reasoning, written communication, civic competences, and English (ICFES, 2019).

As an illustrative case of the use of RBFNNs, the present study will predict the outcomes in the quantitative reasoning section of the SABER PRO test in a cohort of 162,030 Colombian students. The quantitative reasoning section of SABER PRO test measures competencies such as interpreting and representing data, solving problems using mathematical or statistical methods, and obtaining answers and verifying them (ICFES, 2019). Notably, higher education and the workforce have recognized the importance of developing quantitative reasoning skills in the students (Roohr, Graf & Liu, 2014).

The predictors of academic performance selected for this study are well known in educational research. Such a selection strategy makes it possible to exclude redundant predictors aiming to obtain a more valid predictive model. Firstly, *prior academic achievement* has been reported as one of the most important predictors of future academic performance at university (Li, Chen & Duanmu, 2010; McKenzie & Schweitzer, 2001). Secondly, *students' socio-economic status (SES)* has been suggested to be a predictor of academic performance and not merely a covariate (Walpole, 2003). Finally, the predictive role of *high school characteristics* (Black,

Lincove, Cullinane, & Veron, 2015; Wolniak & Engberg, 2010) and *working status* (Salamonson, Everett, Koch, Andrew, & Davidson, 2012; Triventi, 2014) has also been investigated when predicting academic performance in higher education.

Arising out of the preceding, the following research questions are posed in the present study:

1. How to design RBFNNs in order to best predict academic performance in higher education?
2. What is the predictive importance of prior academic achievement, socio-economic conditions, high-school characteristics, and working status when selected as predictors of academic performance in higher education?

Methodology

Description of the Participants

The Colombian Institute for Educational Evaluation (*Instituto Colombiano para la Evaluación de la Educación*, ICFES) provided the data for this study. The total sample included 162,030 students of both genders, females (60%) and males (40%), with a mean age of 23.5 years (*SD* =2.89 years), from private (64.4%) and public universities (35.6%) in Colombia. The students graduated from academic secondary schools (61.6%), both academic and technical secondary schools (18.1%), technical secondary schools (16.7%), and teacher-training secondary schools (3.6%). Furthermore, the students in the sample were enrolled in one of several different academic programs: agronomy, veterinary, and related degrees (1.3%), arts (3.5%), economics, management, and accounting (25.4%), education (8.7%), engineering, architecture, urban planning and related degrees (25.6%), health (11.4%), mathematics and natural sciences (2.1%), and social and human sciences (21%).

Measures

The predictors analyzed in the present study are organized under the measures described next, based on the similarities among the predictors. Descriptive statistics for each predictor are presented in Annex 1.

Prior academic achievement. Students' scores in seven subject areas (biology, physics, chemistry, mathematics, Spanish, social sciences, and philosophy) of the SABER 11 test (between the years 2006 and 2011) are selected as the operationalization of prior academic achievement.

Tuition fees. How students pay for their tuition fees describes this measure. Students pay their tuition fees through either one or several funding sources: parents' resources, educational loans, own resources, and scholarships.

Students' socio-economic status. Parents' educational level, parents' occupation, and monthly family income define students' socio-economic status (Sirin 2005; Van Ewijk and Sleegers 2010).

Students' home characteristics. The number of rooms, whether a student has a computer at home, internet access at home, type of accommodation (permanent or temporal), and home stratum are combined under this measure**.** Home stratum refers to the seven categories used by the Colombian government to classify households based on their physical characteristics and surroundings. The objective of such a classification is to establish the price of public services hierarchically in each area (The World Bank, 2012).

Students' household status. This measure indicates whether a student is head of household and the number of persons the student has in his or her care.

Students' background information. Students' age when taking the SABER PRO test and students' gender define their background information.

High school characteristics. Several high school characteristics are part of this measure. These characteristics are high school academic calendar (indicated by the starting month of the academic year), high school "gender" (boys-only, girls-only, mixed), high school type (public or private), high school schedule (morning-only, afternoon-only, evening-only, weekend-only, ordinary), and high school diploma (academic, technical, academic-technical or teacher-training).

Working status. The number of weekly working hours is the indicator of students' working status (Beauchamp, Martineau, and Gagnon 2016; Triventi 2014).

University background. The number of academic semesters completed by the student when taking the SABER PRO test, university type (public or private), and students' academic program, are sorted under this measure of university background.

Academic performance in higher education. Students' outcomes in the quantitative section of the SABER PRO test from the 2016 administration are chosen as the operationalization of academic performance in higher education. This choice of outcome data is due entirely to the information provided by the ICFES that could be successfully merged with the measures of prior academic achievement, SES, high school characteristics, and working status described above.

Analyses

Preparation of the data

Four steps were carried out to prepare the data analyzed in this study. In step 1, the 12 predictors measured with a continuous scale were standardized (using z-scores), which made comparable their different scales. Furthermore, standardization allowed to presume that all the variables are of the same importance (Garson, 2014). In step 2, the dependent variable was also standardized (using z-scores). In step 3, the 22 categorical predictors were dummy coded. As such, each continuous variable, as well as each category from nominal and categorical variables,

represented an input unit for the RBFNNs. Therefore, 122 units were defined for the input layer of the RBFNNs. In step 4, the total data set (N=162,030) was divided into training data set (80%, n=129,624) and testing data set (20%, n=32,406). Such proportion has previously been recommended in the machine-learning literature (e.g., Suthaharan, 2016).

Designing predictive models based on RBFNNs

SPSS 21 was the software used in this study. SPSS offers a user-friendly module for designing RBFNNs, which permit the modification of the parameters of the network in an easy way and provides the result of the design process through tables and graphs.

To answer the first research question addressed in this study, the design of the predictive models based on RBFNNs comprised two stages: training and testing. In the *training stage*, some parameters of the hidden layer of the RBFNNs were adjusted through four steps. In step 1, the activation function in the hidden layer was set to *ordinary radial basis function*. This function is a Gaussian "bump" as a function of the inputs. In step 2, the number of units in the hidden layer was set to 100, 200, 300, 400, 500, 600, 700, 800, 900, 1k, 5k, 10k, and 25k. It is important to note that the center and the width of the radial basis functions were calculated using the training algorithm specified in Annex 4. The first two steps of the *training stage* led to the training of 13 RBFNNs. In step 3, the accuracy of each trained RBFNN was determined by calculating the correlation coefficient between the expected outcome and the predicted value as well as the relative error. In step 4, the overlapping factor of the 'most' accurate model (identified in step 3) was changed to determine whether there was an increase in the accuracy of the model. SPSS calculates an overlapping default factor given by $1+0.1d$, where d is the number of inputs in the hidden layer of the RBFNNs (IBM, 2013). As the resulting overlapping default factor was equal to 13.2, values above and below 13.2 were chosen for designing the RBFNNs. Therefore, the overlapping factor

was set to 5, 10, 13.2, 20, 40, 80, 100, 200, and 300. The outcome of the *training stage* was the identification of the model, which provided the 'most' accurate prediction.

In the *testing stage*, the 'most' accurate model identified during training was evaluated using the testing data set. The accuracy during the *testing stage* was also verified by calculating the correlation coefficient between the expected and the predicted output of the model as well as the relative error.

Analysis of the predictors of academic performance in higher education

To address the second research question posed in this study, a sensitivity analysis of the selected predictors was carried out for determining their importance in the prediction of academic performance in higher education. The algorithm used for the sensitivity analysis is described in Annex 4. The model used for the analysis was the 'most' accurate model resulting from the design of the RBFNNs. The sensitivity analysis allowed the identification of both the actual importance (as a percentage of a predictive weight) and the normalized importance (percentage with respect to the most important one) of every single predictor. Subsequently, the individual predictors were grouped using the categories described in the section *Measures*. Therefore, it was possible to interpret the predictive contribution of each category of predictors.

Results

The results are given under the research questions addressed in this study.

The design of predictive models based on RBFNNs

Table 1 provides information on the accuracy of each one of the RBFNNs designed in the study. Notice that the number of units in the hidden layer ranged between 100 and 25k. Remarkably, a *saturation* in both the correlation coefficient and the relative error was identified at

the same number of units. In other words, when the number of units was equal to 300, the correlation coefficient was the largest, and the relative error was the lowest, regardless of the subsequent increase in the number of units. This fact is in line with the reasoning behind the design of the RBFNNs, where it has been suggested that the number of units in the hidden layer may be lesser than the number of cases in the training set (Haykin, 2009). As such, the model with 300 units showed the 'most' accurate predictions (r = .594, p = .000, relative error = .647).

The *overlapping factor* among the units in the hidden layer was also changed during the training stage. The reason for this exploration of changes in the overlapping factor was to determine whether this modification influenced the accuracy (i.e., the correlation between the predicted and the observed value as well as the relative error) of the resulting prediction of the 'most' accurate model. Table 2 presents the different values of the overlapping factor analyzed during the training stage. According to Table 2, when the overlapping factor was equal to 5, the correlation of the prediction decreased from .594 to .226, while the relative error increased from .647 to .949. When the overlapping factor was equal to 20, the correlation of the prediction raised to .648, and the relative error decreased to .581. For the subsequent increases in the overlapping factor, there were improvements in both the correlation and the relative error, as shown in Table 2. Once the overlapping factor reached the value of 200, the correlation of the prediction was .701, while the relative error was equal to .509. However, no further improvements were found for larger values of the overlapping factor. Summing up the results of the training stage, the 'most' accurate model was obtained when the *number of units* was equal to 300, and the *overlapping factor* was equal to 200.

In the *testing stage*, the 'most' accurate model was evaluated using the testing data set. Findings suggested that the accuracy of the prediction was satisfactory (r = .704, p = .000, relative error = .504).

The analysis of the predictors of academic performance in higher education

A sensitivity analysis of the 'most' accurate model was carried out to establish the predictive importance of the selected predictors of academic performance in higher education. Table 3 presents the importance as well as the normalized importance of each predictor. On the one hand, the most important single predictors were the scores in three areas of the SABER 11 test: mathematics, chemistry, and biology. Interestingly, the age of the students was found to be the fourth most important single predictor. On the other hand, the least important single predictors were tuition fees coming from a loan, whether the student was head of the household, and tuition fees coming from the parents.

Subsequently, single predictors were grouped using the categories described in the section *Measures*. By doing so, a more meaningful interpretation of the predictive contribution of the selected predictors was achieved. Figure 2 shows the importance of each of the proposed categories of predictors. According to Figure 2, the most important category (63.5%) was *prior academic achievement*. The combination of *socio-economic status* (*SES)* categories at both the individual (i.e., parents' SES, student's household status, and tuition fees) and area levels (i.e., home characteristics) represented the second most important category (14.9%). The third most important category (11.2%) was *background information*. The fourth most important category (7.1%) was *university background*. Remarkably, both *high school characteristics* (2.5%) and *working status* (0.8%) did not show a large predictive contribution to the prediction of academic performance in higher education.

Discussion

This study was carried out to show the potential of RBFNNs as a predictive tool for educational data. The first objective of this study was to design RBFNNs and explore their possible architectures to best predict the academic performance of a cohort of 162,030 Colombian university students. The second objective of this study was to establish the predictive importance of prior academic achievement, socio-economic status, high-school characteristics, and working status when selected as predictors.

Conclusions and implications for educational research

The first conclusion of this study is that several shortcomings of regression analysis could be addressed using RBFNNs. More specifically, the interpretation of categorical predictors, the unfulfillment of several statistical assumptions, and the multicollinearity among the predictors are not major concerns in the case of the RBFNNs. The second conclusion of this study is that sensitivity analysis is a clear and transparent way to determine the predictive contribution of the independent variables in predictive models based on RBFNNs. This fact is particularly relevant when analyzing large data sets.

However, findings of this study do not suggest the dismissal of traditional methodological approaches such as MLR when analyzing educational data. As such, the added value of using the RBFNNs over MLR lies on the interpretability of the resulting predictions. A predictive model based on MLR would result in a regression equation with a high number of regression coefficients given the high number of selected predictors. A clear understanding of such a regression equation would not be easily achieved because of this high number of regression coefficients. Moreover, as many of those regression coefficients would result from regressing the dependent variable on either

nominal or categorical predictors, their interpretation would impose particular difficulties. Therefore, when the interest is to predict academic performance using a vast volume of data, a suitable way to proceed is to use the MLR as an exploratory analysis of the relationships of interest. Then, more robust approaches such as RBNNs should be considered in the analysis.

Recommendations for designing RBFNNs

Regarding the first research question, the findings of this study are providing specific recommendations to consider when designing predictive systems based on RBFNNs. To begin with, it was found that changes in the hidden layer of the RBFNNs influence their accuracy accordingly to the modified parameter. In the case of the number of units, the results of this study support the idea that it is possible to design accurate RBFNNs with fewer units in the hidden layer than exemplars in the training data set. Moreover, *saturation* in the value of both accuracy measures was reached when a specific number of units was chosen. It could be argued, then, that with less time of training, there are still accurate predictions in the RBFNNs. To determine the 'best' number of units, however, represents a challenging task for the design of the RBFNNs. Although SPSS 21 includes a feature to calculate the 'best' number of units based on Bayesian information criterion (BIC) and Akaike information criterion (AIC), more accurate predictions were found after systematically determining the number of units. What seems to work better, then, is to increase the number of units gradually and to compare its effect on the accuracy of the prediction. In this way, the drawbacks of carrying out a trial and error process can be minimized.

Furthermore, this study also revealed that the overlapping factor among the units does cause changes in the accuracy of the resulting predictions. However, there are no specific rules to calculate such overlapping factor. Once again, the procedure followed in this study suggests to systematically modify the overlapping factor to determine changes in the accuracy of the

prediction. Notably, it was found that changes in the overlapping factor among the units produced changes of lesser value in the accuracy of the RBFNNs than changes in the number of units. Hence, a recommendation for the design of the RBFNNs is to set first the number of units and, next, adjust the overlapping factor among the units until reaching a satisfactory level of accuracy in the prediction.

Finally, it is essential to highlight the fact that the training of the RBFNNs involves two distinct phases. First, the centers and scaling parameters of the hidden layer are calculated through unsupervised or supervised learning. Second, the weights of the output layer are calculated through supervised learning (Schwenker et al., 2002). Therefore, it is advisable to explore how several different training algorithms influence the performance of RBFNNs and the accuracy of the prediction.

Remarks for analyzing academic performance in higher education

Regarding the second research question, the findings of this study can be transferred into three main remarks to consider when predicting academic performance in higher education. The first remark is the existence of both direct and indirect influence of SES on academic performance in higher education. This study showed that prior academic achievement is a strong predictor of future academic performance in university, which corroborates previous predictive studies in higher education. However, it was also found that individual and area-level SES indicators were predictors of academic performance, despite the SES's influence on prior academic achievement. Hence, any predictive model, including SES as a predictor, should also simultaneously investigate the predictive role of prior academic achievement.

The second remark is that several other predictors could mask the influence of high school characteristics on academic performance in higher education. More specifically, it could be the case that prior academic achievement is already capturing any predictive role of high school characteristics. It can be argued that students coming from high schools with better conditions also show better prior academic achievement before university. Moreover, it could be the case that SES also captures part of the influence of high school on academic performance in university. As such, students coming from most privileged families could also attend high schools with better infrastructure and more qualified teachers.

The third remark is the need for considering the students' academic program when predicting their academic performance in university. As such, the analysis carried out in this study showed that students' academic program provided valuable information to the prediction of academic performance. Undoubtedly, students choose to study an academic program based on their interests and skills. However, it could be the case that differences across the academic programs about content knowledge, assessment practices, and learning experiences cause that students' academic performance develops differently. Therefore, the influence of the predictors of academic performance might also vary among different fields of study.

Limitations

Technical and educational limitations of this study warrant specific mention. Regarding the technical limitations, only one type of Gaussian function was selected for designing the RBFNNs. However, there are additional types of Gaussian function which could show better performance but were not considered in the analysis. Furthermore, due to the absence of a validation set, it was not possible to indeed generalize the results of the tested RBFNNs models. Finally, all the RBFNNs models analyzed in this study used radial basis functions with a constant

width. However, it is rather important to adequately select such parameters, which also influences the performance of the RBFNNs models (Benoudjit & Verleysen, 2003).

Regarding the educational limitations, no students' cognitive factors were included as predictors of academic performance. The SABER 11 test was the only indicator of students' prior academic achievement. Furthermore, given the sample size analyzed in this study (N = 162,030), it was not possible to include more traditional indicators of academic performance in higher education as the grade point average (GPA), not available in the data set provided. Finally, only one attribute of the students' working status was considered for analysis (number of hours worked). No information on reasons for working nor type of work (i.e., outside or inside the academic field) could be retrieved for the analysis. Hence, the predictive contribution of working status could not be fully determined.

Future research

Future work on the use of RBFNNs in educational research could investigate both methodological and educational topics. In relation to the methodological topics, different algorithms for the training of the RBFNNs should be explored. In particular, the weights in the output layer of the RBFNNs could be calculated utilizing supervised algorithms, different from the ordinary least squares. For instance, clustering, vector quantization, or classification tree algorithms (Schwenker et al., 2002).

Moreover, it would be of great interest to consider the use of different Gaussian functions as the activation function for the units in the hidden layer of the RBFNNs. Finally, the design of multilayer RBFNNs (i.e., RBFNNs with multiple hidden layers) could be a relevant topic of special interest when using machine-learning techniques in the educational research field. Regarding

educational issues, higher education institutes could develop early detection systems of academic performance based on RBFNNs, so that the students who are at risk of failure can be identified. In this way, the intervention strategies could be efficiently focused on that group of students.

References

Alin, A. (2010). Multicollinearity. *Wiley Interdisciplinary Reviews: Computational Statistics*, *2*, 370-374. doi:10.1002/wics.84

Areerachakul, S., & Sanguansintukul, S. (2009). A comparison between the multiple linear regression model and neural networks for biochemical oxygen demand estimations. *Proceedings of the Eighth International Symposium on Natural Language Processing*, 11-15. doi:10.1109/snlp.2009.5340937

Beauchamp, G., Martineau, M., & Gagnon, A. (2016). Examining the link between adult attachment style, employment and academic achievement in first semester higher education. *Social Psychology of Education, 19,* 367-384. doi:10.1007/s11218-015-9329-3

Benoudjit, N., & Verleysen, M. (2003). On the kernel widths in radial-basis function networks. *Neural Processing Letters, 18,* 139-154. doi:10.1023/A:1026289910256

Berens, J., Schneider, K., Görtz, S., Oster, S., & Burghoff, J. (2018). *Early detection of students at risk–Predicting student dropouts using administrative student data and machine learning methods* (Schumpeter Discussion Papers No. sdp18006). Retrieved from EconPapers website: https://econpapers.repec.org/paper/bwuschdps/sdp18006.htm

Black, S. E., Lincove, J., Cullinane, J., & Veron, R. (2015). Can you leave high school behind? *Economics of Education Review*, *46,* 52-63. doi:10.3386/w19842

Bonsaksen, T. (2016). Predictors of academic performance and education programme satisfaction in occupational therapy students. *British Journal of Occupational Therapy*, *79,* 361-367. doi:10.1177/0308022615627174

Broomhead, D. S., & Lowe, D. (1988). Multivariable functional interpolation and adaptive networks. *Complex Systems, 2,* 321-355.

Chan, Y.H. (2004). Biostatistics 201: Linear regression analysis. *Singapore Medical Journal, 45,* 55-61.

Cohn, E., Cohn, S., Balch, D. C., & Bradley Jr, J. (2004). Determinants of undergraduate GPAs: SAT scores, high-school GPA and high-school rank. *Economics of Education Review, 23*, 577-586. doi: 10.1016/j.econedurev.2004.01.001

Dash, C. S. K., Behera, A. K., Dehuri, S., & Cho, S. B. (2016). Radial basis function neural networks: a topical state-of-the-art survey. *Open Computer Science, 6,* 33-63. doi:10.1515/comp-2016-0005

De Clercq, M., Galand, B., Dupont, S., & Frenay, M. (2013). Achievement among first-year university students: An integrated and contextualized approach. *European Journal of Psychology of Education, 28,* 641-662. doi:10.1007/s10212-012-0133-6

DeShields Jr, O. W., Kara, A., & Kaynak, E. (2005). Determinants of business student satisfaction and retention in higher education: applying Herzberg's two-factor theory. *International Journal of Educational Management, 19,* 128-139. doi: 10.1108/09513540510582426

Ever, Y. K., & Dimililer, K. (2018). The effectiveness of a new classification system in higher education as a new e-learning tool. *Quality and Quantity,52*, 573-582. doi: 10.1007/s11135-017-0636-y

Fausett, L. V. (1994). *Fundamentals of neural networks*. New Jersey, USA: Prentice-Hall.

Field, A. P. (2009). *Discovering statistics using SPSS: (and sex and drugs and rock 'n' roll).* Los Angeles, California: SAGE Publications.

Garson, G. D. (2014). *Neural network models*. Asheboro, NC: Statistical Associates Publishers.

Gerken, J. T., & Volkwein, J. F. (2000, May). *Pre-college characteristics and freshman year experiences as predictors of 8-year college outcomes*. Paper presented at the annual meeting of the Association for Institutional Research, Cincinnati, OH.

Haviluddin, H., Sunarto, A., & Yuniarti, S. (2014). A Comparison between simple linear regression and radial basis function neural network (RBFNN) models for predicting students' achievement. *Proceedings of the International Conference on Education (ICEdu14)*, 299-308. doi: 10.13140/2.1.3878.5600

Haykin, S. S. (2009). *Neural networks and learning machines* (Vol. 3). Upper Saddle River, NJ, USA: Pearson.

Huang, S., & Fang, N. (2013). Predicting student academic performance in an engineering dynamics course: A comparison of four types of predictive mathematical models. *Computers & Education, 61,* 133-145. doi: 10.1016/j.compedu.2012.08.015

IBM. (2013). *IBM SPSS neural networks 22*. Armonk, NY: IBM.

IBM. (2019, February 27). Output Statistics (radial basis function algorithms). Retrieved from http://desktop-5vfq00m:55284/help/index.jsp?topic=/com.ibm.spss.statistics.help/idh_idd_rbf_variables.htm

ICFES (2019). Información de la prueba SABER PRO. [Specifications about the SABER PRO test]. Retrieved from: http://www.icfes.gov.co/instituciones-educativas-y- secretarias/saber-pro/informacion-de-la-prueba-saber-pro

Jeatrakul, P., & Wong, K. W. (2009). Comparing the performance of different neural networks for binary classification problems. *Proceedings of the Eighth International Symposium on Natural Language Processing*, 111-115. doi: 10.1109/snlp.2009.5340935

Jeon, J. (2015). The strengths and limitations of the statistical modeling of complex social phenomenon: Focusing on SEM, path analysis, or multiple regression models. *International Journal of Social, Behavioral, Economic, Business and Industrial Engineering*, *9*, 1634-1642.

Kara, A., & DeShields Jr, O. W. (2004). Business student satisfaction, intentions and retention in higher education: An empirical investigation. *Marketing Educator Quarterly*, *3*, 1-25.

Kayri, M., & Çokluk, Ö. (2016). Examining factors of academic procrastination tendency of university students by using artificial neural network. *International Journal of Computer Trends and Technology*, *34*, 1-8. doi: 10.14445/22312803/ijctt-v34p101

Kent, R. (2009). Rethinking data analysis- part two. Some alternatives to frequentist approaches. *International Journal of Market Research, 51,* 181–202. doi:10.2501/s1470785308200304

Kriesel, D. (2005). *A brief introduction to neural networks*. Retrieved from: http://www.dkriesel.com/_media/science/neuronalenetze-en-zeta2-2col-dkrieselcom.pdf

Kyndt, E., Musso, M, Cascallar, E. & Dochy, F. (2015). Predicting academic performance: The role of cognition, motivation and learning approaches. A neural network analysis. In V. Donche, S. De Maeyer, D. Gijbels, & H. van den Bergh (Eds). *Methodological challenges in research on student learning*, (pp. 55-76) Garant: Antwerp, Belgium

Lau, E. T., Sun, L., & Yang, Q. (2019). Modelling, prediction and classification of student academic performance using artificial neural networks. *SN Applied Sciences*, *1*, 982. doi: /10.1007/s42452-019-0884-7

Li, G., Chen, W., & Duanmu, J. L. (2010). Determinants of international students' academic performance: A comparison between Chinese and other international students. *Journal of Studies in International Education, 14*, 389-405. doi:10.1177/1028315309331490

Marshall, David B., English, D. J. (2000). Neural network modeling of risk assessment in child protective services. *Psychological Methods*, *5*, 102–124. doi: 10.1037//1082-989x.5.1.102

McKenzie, K., & Schweitzer, R. (2001). Who succeeds at university? Factors predicting academic performance in first year Australian university students. *Higher Education Research and Development*, *20*, 21-33. doi:10.1080/07924360120043621

Muijs, D. (2004). *Doing quantitative research in education with SPSS*. London, UK: Sage.

Musso, M., Kyndt, E., Cascallar, E., & Dochy, F. (2012). Predicting mathematical performance: The effects of cognitive processes and self-regulation factors. *Education Research International*, 2012, 1-13, doi:10.1155/2012/250719.

Musso, M. F., Kyndt, E., Cascallar, E. C. & Dochy F. (2013). Predicting general academic performance and identifying the differential contribution of participating variables using artificial neural networks. *Frontline Learning Research*, *1*, 42-71, doi:10.14786/flr.v1i1.13.

Nimon, K. F. (2012). Statistical assumptions of substantive analyses across the general linear model: a mini review. *Frontiers in Psychology*, *3*, 1-5. doi:10.3389/fpsyg.2012.00322

Orr, M. J. (1996). *Introduction to radial basis function networks*. Retrieved from: http://www.cc.gatech.edu/~isbell/tutorials/rbf-intro.pdf

Osborne, J., & Waters, E. (2002). Four assumptions of multiple regression that researchers should always test. *Practical Assessment, Research & Evaluation*, *8*, 1-5.

Paik, H. (2000). Television viewing and high school mathematics achievement: A neural network analysis. *Quality and Quantity*, *34*, 1-15. doi:10.1023/a:1004795407624

Paliwal, M. and Kumar, U.A. (2009). A study of academic performance of business school graduates using neural network and statistical techniques. *Expert Systems with Applications*, *36*, 7865-7872. doi: 10.1016/j.eswa.2008.11.003

Pao, H. T. (2008). A comparison of neural network and multiple regression analysis in modelling capital structure. *Expert Systems with Applications, 35,* 720-727. doi: 10.1016/j.eswa.2007.07.018

Roohr, K. C., Graf, E. A., & Liu, O. L. (2014). Assessing quantitative literacy in higher education: An overview of existing research and assessments with recommendations for next-generation assessment. *ETS Research Report Series, 2,* 1-26.

Rothstein, J. M. (2004). College performance predictions and the SAT. *Journal of Econometrics, 121,* 297-317. doi: 10.1016/j.jeconom.2003.10.003

Sadler, D. R. (2013). Making competent judgments of competence. In Blömeke S., Zlatkin-Troitschanskaia O., Kuhn C., Fege J. (Eds.), *Professional and Vet Learning: Vol. 1. Modelling and Measuring Competencies in Higher Education* (pp. 13-27). Dordrecht, The Netherlands: Sense Publishers. doi:10.1007/978-94-6091-867-4

Salamonson, Y., Everett, B., Koch, J., Andrew, S., & Davidson, P. M. (2012). The impact of term-time paid work on academic performance in nursing students: A longitudinal study. *International Journal of Nursing Studies, 49,* 579-585. doi: 10.1016/j.ijnurstu.2011.10.012

Schwenker, F., Kestler, H. A., & Palm, G. (2002). Unsupervised and supervised learning in radial-basis-function networks. In Seiffert U., Jain L.C. (Eds), *Studies in Fuzziness and Soft Computing: Vol. 78. Self-organizing neural networks* (pp. 217-243). Heidelberg: Physica-Verlag. doi: 10.1007/978-3-7908-1810-9_10

Shahiri, A. M., Husain, W., & Rashid, N.A. (2015). A review on predicting student's performance using data mining techniques. *Procedia Computer Science, 72,* 414-422. doi: 10.1016/j.procs.2015.12.157

Shavelson, R. J. (2010). On the measurement of competency. *Empirical Research in Vocational Education and Training, 2*, 41-63.

Sirin, S. (2005). Socioeconomic status and academic achievement: A meta-analytic review of research. *Review of Educational Research, 75*, 417-453. doi:10.3102/00346543075003417

Spiess, A. N., & Neumeyer, N. (2010). An evaluation of R^2 as an inadequate measure for nonlinear models in pharmacological and biochemical research: a Monte Carlo approach. *BMC Pharmacology, 10,* 6. doi: 10.1186/1471-2210-10-6

Suthaharan S. (2016). Supervised Learning Algorithms. In: R. Sharda & S. Voß (Eds.), *Integrated Series in Information Systems: Vol.36. Machine learning models and algorithms for big data classification* (pp. 183-206). Boston, MA: Springer.

Tacq, J. (1997). *Multivariate analysis techniques in social science research: From problem to analysis.* London, UK: Sage.

Teshnizi, S. H., & Ayatollahi, S. M. T. (2015). A comparison of logistic regression model and artificial neural networks in predicting of student's academic failure. *Acta Informatica Medica, 23*, 296–300, doi: 10.5455/aim.2015.23.296-300

The World Bank (2012). *Reviews of national policies for education: Tertiary education in Colombia 2012.* Paris: OECD Publishing. doi:10.1787/9789264180697-en

Triventi, M. (2014). Does working during higher education affect students' academic progression? *Economics of Education Review, 41,* 1-13. doi:10.1016/j.econedurev.2014.03.006

Van Ewijk, R., & Sleegers, P. (2010). The effect of peer socioeconomic status on student achievement: A meta-analysis. *Educational Research Review, 5,* 134-150. doi:10.1016/j.edurev.2010.02.001

Walpole, M. (2003). Socioeconomic status and college: How SES affects college experiences and outcomes. *The review of Higher Education, 27,* 45-73. doi:10.1353/rhe.2003.0044

Wang, H., Kong, M., Shan, W., & Vong, S. K. (2010). The effects of doing part-time jobs on college student academic performance and social life in a Chinese society. *Journal of Education and Work, 23,* 79-94. doi:10.1080/13639080903418402

Webber, D. A., & Ehrenberg, R. G. (2010). Do expenditures other than instructional expenditures affect graduation and persistence rates in American higher education? *Economics of Education Review, 29,* 947-958. doi: 10.3386/w15216

Williams, M., Gómez, C., & Kurkiewicz, D. (2013). Assumptions of multiple regression: correcting two misconceptions. *Practical Assessment, Research & Evaluation, 18,* 1-14.

Wolniak, G. C., & Engberg, M. E. (2010). Academic achievement in the first year of college: Evidence of the pervasive effects of the high school context. *Research in Higher Education, 51,* 451-467. doi:10.1007/s11162-010-9165-4

York, T. T., Gibson, C., & Rankin, S. (2015). Defining and measuring academic success. *Practical Assessment, Research & Evaluation, 20,* 1-20.

Zheng, J. L., Saunders, K. P., Shelley, I. I., Mack, C., & Whalen, D. F. (2002). Predictors of academic success for freshmen residence hall students. *Journal of College Student Development, 43,* 267-283.

Table 1. Training of the RBFNNs

Number of units in the hidden layer	*r coefficient for ordinal basis function*	*Relative error*
100	.538*	.712
200	.572*	.673
300	.594*	.647
400	.594*	.647
500	.594*	.647
600	.594*	.647
700	.594*	.647
800	.594*	.647
900	.594*	.647
1000	.594*	.647
5000	.594*	.647
10000	.594*	.647
25000	.594*	.647

*Note. * denotes a significant correlation at .000 level.*

Table 2. Training results modifying the overlapping factor (n=300)

Overlapping factor	*r coefficient*	*Relative error*
5	.226*	.949
10	.538*	.711
13.2	.594*	.647
20	.648*	.581
40	.686*	.530
80	.698*	.514
100	.699*	.512
200	.701*	.509
300	.701*	.509

*Note. * denotes a significant correlation at .000 level.*

Table 3. Predictive importance of the selected predictors

Predictor	*Importance*	*Normalized Importance*
SABER 11-mathematics	.188	100.0%
SABER11-chemistry	.112	59.5%
SABER11-biology	.112	59.4%
Age at SABERPRO	.092	49.0%
SABER11-social sciences	.082	43.8%
SABER 11-spanish	.061	32.3%
SABER11-physics	.057	30.0%
Academic program	.040	21.3%
Number of persons in charge	.039	20.9%
SABER11-philosophy	.023	12.0%
Student's gender	.020	10.4%
Monthly family income	.018	9.3%
Number of rooms	.016	8.6%
University type	.016	8.3%
Number of semesters at SABER PRO	.015	8.2%
Mother's educational level	.012	6.2%
Students' house stratum	.011	5.6%
Father's occupation	.010	5.3%
Father's educational level	.010	5.2%
Mother's occupation	.009	4.7%
High school schedule	.008	4.4%
Number of hours worked	.008	4.3%
Ownership of a PC	.007	3.8%
High school academic calendar	.006	3.1%
Tuition fees (scholarship)	.005	2.6%
High school diploma	.004	2.3%
High school "gender"	.004	2.1%
Internet access	.003	1.4%
Tuition fees (own resources)	.002	1.2%
High school type (public or private)	.002	1.2%
Type of house	.002	1.1%
Tuition fees (loan)	.002	1.1%
Head of household	.002	1.1%
Tuition fees (parents' resources)	.002	1.0%

Figure 1. Architecture of a radial basis function neural network

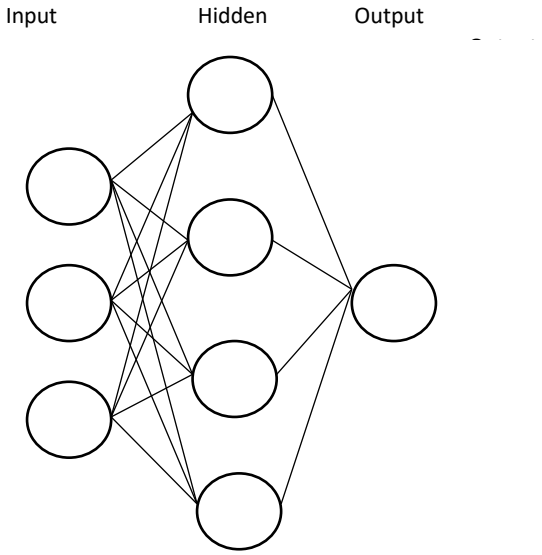
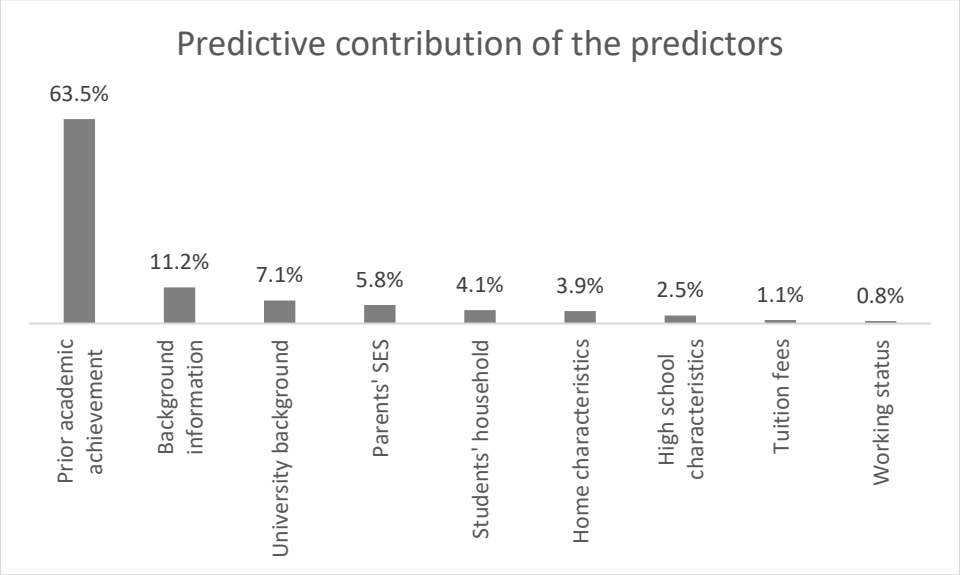


Figure 2. Predictive contribution of the predictors of academic performance in higher education



Chapter 5

A cluster analysis of academic performance in higher education through self-organizing maps (SOMs)

Abstract

Cluster analysis is a powerful tool to explore educational data in order to create groups of students, teachers, or institutions according to their similarities. Self-organizing maps (SOMs) is one of the existing clustering methods, which makes it possible to cluster large data sets, to identify hidden patterns in the data, and to visualize complex relationships among the investigated variables. Interestingly, SOMs find relationships in data sets where information on the output is unknown. Therefore, it seems worthy to further extend the use of SOMs to cluster educational data. This study applies self-organizing maps (SOMs) to a cohort of 162,030 university students to determine clusters based on their prior academic achievement, socio-economic status (SES), and academic performance in higher education. Findings suggest that interesting patterns exist between prior academic achievement, academic performance in higher education, and SES. Starting from these variables, students can be clustered into three groups, which are statistically different among them.

Keywords: self-organizing maps, cluster analysis, academic performance, higher education.

Introduction

Recent developments in computer science have made it possible to collect and analyze data of increasing size and high dimensionality (Wehrens & Buydens, 2007). In line with these recent developments, the emerging field of educational data mining (EDM) has aimed at extracting useful information from educational data sets to positively influence the educational field (Romero & Ventura, 2010). Among several other research interests, the analysis of students' academic performance is considered a core topic within EDM (Baker, 2010; Gibson & Ifenthaler, 2017; Romero & Ventura, 2010; Shahiri, Husain, & Rashid, 2015). EDM facilitates the development of predictive models that can accurately estimate the present and future states of students' academic performance (Daniel, 2017).

Within the variety of techniques utilized in EDM, cluster analysis is the one that deals with the problem of partitioning educational data into subsets (Dutt, Aghabozrgi, Ismail, & Mahroeian, 2015). In the resulting partitions, each subset should include objects which are similar among them, and each subset should be statistically different from the others (Huberty, Jordan, & Brandt, 2005). Cluster analysis also identifies which variables contribute the most to separate the created subsets (Antonenko, Toy, & Niederhauser, 2012).

The use of cluster analysis has not gone unnoticed among higher education researchers, as their interest has been to determine whether students, teachers or institutions may be grouped in relation to their perceptions, attitudes, and other characteristics (Huberty et al., 2005). Several clustering methods, such as k-means clustering, hierarchical clustering, and self-organizing maps (SOMSs), have been applied in past educational research to analyze students' academic performance (Kausar, Huahu, Hussain, Wenhao, & Zahid, 2018). Two essential conclusions might be drawn from this body of research.

On the one hand, it has been suggested that prior academic achievement and socio-economic status (SES) are influential variables to create not only students' subgroups but also contribute to differentiate their academic performance in higher education (De Clercq, Galand, & Frenay, 2017; Da Silva, de Fátima Nunes, Santos, Queiros, & Leles, 2012; Rodríguez, 2016). On the other hand, there is no definitive evidence on which clustering method shows the best performance, as each one uses different algorithms as well as various similarity measures to group the data (Kumar & Dhamija, 2010). Nevertheless, Antonenko et al. (2012) have suggested that k-means and hierarchical clustering might be impractical to analyze large data sizes. As such, k-means clustering is computationally complex when processing a vast volume of information (Qi, Yu, Wang, Liu, & Wang, 2017). Moreover, neither k-means nor hierarchical clustering provides clear insights into the relationship among the input variables in the resulting grouping.

Therefore, when the interest is to discover existing patterns in large data sets, the most appropriate clustering method seems to be self-organizing maps (SOMs) for the reasons presented next. First, SOMs are "self-organized" models capable of identifying unknown patterns in large data sets where no input-output relationship is specified in the data. Second, SOMs are "maps" that place any multidimensional input data (i.e., several variables and thousands of cases) in a two-dimensional plane, that it is easier to interpret and understand. Third, SOMs also provide a graphical representation of the way in which the variables in the input data are related. Starting from these features of SOMs, it seems worthwhile to explore whether their use contributes to a better understanding of students' academic performance in higher education.

Given these considerations, the objective of this study is twofold. The first objective is to analyze the existing patterns between prior academic achievement, SES, and academic performance in a cohort of 162,030 Colombian university students through a SOM model. The

measure of academic performance is the outcome in the quantitative reasoning section of the SABER PRO test, a compulsory standardized test developed and administered to university students in Colombia. The second objective is to cluster the resulting SOM to determine specific groups of students based on their prior academic achievement, SES, and academic performance in higher education. The expected contribution of this study is to extend the use of SOMs for the analysis of educational data sets.

State of the art

This section provides a general description of self-organizing maps (SOMs), including information on its definition, structure, learning paradigm, design parameters, and quality measures. Besides, this section discusses examples of the previous use of SOMs in educational research.

Self-organizing maps (SOMs)

Two types of predictive models based on artificial neural networks can be distinguished: supervised and unsupervised models. On the one hand, supervised models use patterns in the data to iteratively compare output estimates to observed estimates, while reducing the difference between them. On the other hand, unsupervised models directly search for patterns in the data and group similar cases into clusters (Garson, 2014). As such, there is no information on the output variable in unsupervised models. Self-organizing maps (SOMs) are a type of unsupervised artificial neural network developed by Teuvo Kohonen in 1982. The main goal of SOMs is to transform a high-dimension input data set into a two-dimensional map but maintaining the existing way in which the data is related (Haykin, 2009; Kohonen, 2013). In other words, if two elements

from the data set exhibit similar patterns in the input variables, their position in the resulting SOM would also be very similar (Wehrens & Buydens, 2007).

Structure of SOMs

The typical structure of SOMs is shown in Figure 1. According to Haykin (2009), SOMs are layered structures with two layers. In the *input layer*, the number of neurons is equal to the number of cases in the data set. As in any other artificial neural network, the input layer connects the network to the data set inputs. In the *competitive layer*, the neurons are organized as the nodes of an empty grid, which is usually two dimensional. The size of the grid determines the number of neurons in the competitive layer (e.g., a square grid of dimension five would have 25 neurons). Each neuron of the input layer is linked to every neuron in the competitive layer. Therefore, all the information provided to the input layer is connected to all the neurons in the competitive layer.

Competitive learning of SOMs

SOMs identify existing patterns in large data sets through competitive learning (Haykin, 2009). The objective of this process is to find in the competitive layer, the neuron "most similar" to every neuron in the input layer (Iturriaga & Sanz, 2013), following the steps described below (Haykin, 2009; Kohonen, 2013):

1. Each case in the data set is transformed into a n-dimensional input vector, where n indicates the number of input variables in the data set.
2. A grid of neurons is created. Next, each neuron in the grid is initialized with a random n-dimensional weight vector. This weight vector is usually known as the prototype of that neuron.

3. Every input vector (i.e., each case in the data set) is compared to all the prototypes in the grid. The neurons in the grid compete among themselves to win each of the input vectors. The winner neuron is the one whose prototype is closer to the input vector, and it is known as the best matching unit (BMU).
4. The weight vector of the BMU, as well as the weight vectors of the surrounding neurons, are adjusted to match best with the input vector. This adaptative process is possible because of a so-called *neighborhood function* that is defined around the BMU. Therefore, the most substantial transformations occur in the nearest neurons to the BMU, while the smaller changes happen in the farthest neurons.
5. The updating process of the weight vector is described by Equation 1.

$$w_j\,(n+1) = w_j\,(n) + alfa\,(n)\,h_{j,i(x)}\,(n)\,(\boldsymbol{x}(n) - w_j\,(n)) \quad (1)$$

Each component of Equation 1 is explained subsequently. $w_j\,(n)$ is the weight vector of the *jth* neuron, *alfa* represents the learning rate of the competitive learning, $h_{j,i(x)}$ describes the neighborhood function, and $\boldsymbol{x}$ is the input vector. During the updating, the learning rate as well as the size of the neighborhood decrease.

Designing of SOMs

Several parameters should be tuned when designing SOMs (Kohonen, 2014). First, the form and the size of the grid (i.e., the number of neurons) are defined. Usually, the form of the grid is rectangular. Next, the shape of the neurons in the grid (rectangular or hexagonal) is selected. Such a parameter is rather crucial because it defines the topological relation of the neurons in the SOM. To put it differently, a rectangular neuron would have eight surrounding neurons, while a hexagonal neuron would have six neighboring neurons. In addition, the weight vector of each neuron is initialized using either random initialization (weights are small random values) or linear

initialization (weights are obtained from the space generated by the linear combination of the two principal components of the input data). Finally, the type of neighborhood function (typically Gaussian) and the type of training (batch or sequential) are also set.

SOMs quality measures

There are two traditional ways to evaluate the quality of SOMs: the quantization error and the topographic error (Pölzlbauer, 2004). On the one hand, the quantization error occurs because all the input vectors included in a BMU are similar but not precisely equal (Peiró-Velert et al., 2014). Therefore, a difference exists between each input vector and the prototype of its respective BMU. The quantization error is simply the average of the distances between each input vector and the prototype of its corresponding BMU. Smaller values of quantization error suggest a better fit of the data.

On the other hand, the topographic error refers to how well the obtained map resembles the input data (Pölzlbauer, 2004). The topographic error is calculated by determining first the best matching unit and the second-best matching unit for each input vector and then evaluating their positions. An error occurs if the best matching units are not contiguous in the grid. The topographic error ranges between 0 and 1, where 0 indicates perfect topology preservation.

SOMs for the analysis of academic performance in educational research

SOMs have successfully been used to analyze educational data in recent educational literature. In the study of Pellicer-Chenoll et al. (2015), the relationship between physical activity, physical fitness, and academic achievement of 444 high school students was investigated using SOMs. The authors claimed to select SOMs, as a methodological alternative to the traditional linear analysis, aiming at modeling a non-linear relationship among the selected variables. Pellicer-

Chenoll et al. (2015) found four clusters to group the students and described them using one of two profiles. The first profile corresponded to the students with low physical activity, less physical fitness, and a lower GPA. In contrast, the second profile described the students with high physical activity, higher physical fitness, and a higher GPA.

Besides, Alias, Ahmad, and Hasan (2015) used SOMs to analyze the students' actions while using an e-learning system in a group of 19 undergraduate students in Malaysia, as SOMs were very useful to identify hidden patterns in the data set as well as providing interesting information for future analysis on students' academic performance. The analyzed data set contained 2394 hits resulting from the activity of the students after 14 weeks of academic work. Alias et al. (2015) identified four clusters based on the similarities in students' hits when using the e-learning system. Students belonging to cluster 2 were the most active, showing a high value of similarities in the analyzed browsing activities such as downloading examples on specific topics, course viewing, and assignment viewing and submission.

Finally, Lee (2019) used SOMs to investigate the problem-solving pattern of 4,337 students enrolled in *edX 8.MReVx Mechanics Review*, a massive open online course (MOOC) offered by the Massachusetts Institute of Technology (MIT) in Summer 2014. Lee (2019) reports that SOMs might help to support the learning process of students aiming to increase the completion rate of the course. The problem-solving pattern of the students was represented by 15 variables, which indicated their daily, weekly, and semesterly completion percentage of homework and quiz problems. After finding a SOM model to first group the students, Lee (2019) identify four clusters of students through hierarchical clustering. Cluster 1 includes students who dropped out of the course. Cluster 2 consists of the students who actively solved homework and quiz problems but

left the course after the third week of the semester. Students belonging to clusters 3 and 4 had a high course completion rate (93%) and solved more than 50% of homework and quiz problems.

The present study

The general trend in higher education literature has been to collect data on different predictors of academic performance in order to confirm or refute theoretical claims previously established. Nonetheless, this approach possibly fails at identifying several existing patterns among the data set of predictors. Furthermore, as the size of the data sets containing students' information keeps increasing, the number of unknown patterns is also growing. These patterns are relevant as they could provide critical information about learning styles, effective teaching practices, or positive institutional experiences. Therefore, by identifying such patterns, a better understanding of academic performance in higher education could be achieved.

Cluster analysis emerges as one powerful alternative for pattern recognition because it makes it possible to organize students in specific subgroups according to their characteristics (Romero & Ventura, 2010); and to use the resulting subgroups for subsequent analysis (Antonenko et al. 2012). Among the existing clustering techniques, self-organizing maps (SOMs) are an unsupervised type of artificial neural network which deals with large data sizes, reduces the dimension of any given input space, preserves the relationship among the input variables, and provides a graphical representation of such association (Haykin 2009; Kohohen, 2013).

The objective of the present study is twofold. The first objective is to analyze the existing patterns between prior academic achievement, SES, and academic performance in a cohort of 162,030 Colombian university students through a SOM model. The second objective is to cluster the resulting SOM to determine specific groups of students based on their prior academic

achievement, SES, and academic performance in higher education. Consequently, the research questions posed in the present study are:

1. Which patterns between prior academic achievement, SES, and academic performance in higher education can be identified through a SOM model?
2. How many groups of students exist in the resulting SOM based on their prior academic achievement, SES, and academic performance in higher education?

Methodology

Description of the Participants

The Colombian Institute for Educational Evaluation (*Instituto Colombiano para la Evaluación de la Educación*, ICFES) provided the data for this study. The total sample included 162,030 students of both genders, females (60%) and males (40%), with a mean age of 23.5 years (*SD* =2.89 years), from private (64.4%) and public universities (35.6%) in Colombia. The students graduated from academic secondary schools (61.6%), both academic and technical secondary schools (18.1%), technical secondary schools (16.7%), and teacher-training secondary schools (3.6%). Furthermore, the students in the sample were enrolled in one of several different academic programs: agronomy, veterinary, and related degrees (1.3%), arts (3.5%), economics, management, and accounting (25.4%), education (8.7%), engineering, architecture, urban planning and related degrees (25.6%), health (11.4%), mathematics and natural sciences (2.1%), and social and human sciences (21%).

Measures

The input variables selected for this study are organized under the measures described next. Annex 1 presents the descriptive statistics for each input variable.

Prior academic achievement. Students' scores in seven subject areas (Spanish, mathematics, social sciences, philosophy, biology, chemistry, and physics) of the SABER 11 test (between the years 2006 and 2011) are selected as the operationalization of prior academic achievement. SABER 11 is a standardized test required for admission to higher education institutions in Colombia. This is a cluster of exams that assess students' level of achievement in eight subject areas at the end of high school. Each subject test consists of multiple-choice questions with only one correct answer and three distractors. Test results are reported using a scale from 0 to 100 for each subject, resulting from an item response theory (IRT) analysis of the data.

Students' socio-economic status. Information regarding parents' educational level and monthly family income are selected as indicators of students' socio-economic status (Sirin 2005; Van Ewijk & Sleegers 2010). Parents' educational level was assessed under the International Standard Classification of Education (ISCED). The ISCED is a classification of education programs by levels and areas of knowledge proposed in 1997 by the United Nations Educational, Scientific, and Cultural Organization (UNESCO) and then revised in 2011. There are nine ISCED levels from Early childhood education to Doctoral or equivalent degree. Monthly family income was measured as a multiple of the monthly minimum wage (i.e., one minimum wage, twice minimum wages, etc.).

Students' home stratum. Home stratum refers to the seven categories used by the Colombian government to classify households based on their physical characteristics and surroundings. The main reason behind this classification is to establish the price of public services hierarchically in each area (The World Bank 2012).

Academic performance in higher education. Students' outcomes in the quantitative section of the SABER PRO test (hereafter, SABER PRO test) from the 2016 administration are chosen as the operationalization of academic performance in higher education.

SABER PRO is a standardized test required for graduation from higher education institutions in Colombia. The exam assesses students' performance in the following sections: critical reading, quantitative reasoning, written communication, civic competences and English. Each section of the test consists of multiple-choice questions with only one correct answer and three distractors. Additionally, SABER PRO evaluates specific competences such as scientific thinking for science and engineering students; teaching, training and evaluating for education students; conflict management and judicial communication for law students; agricultural and livestock production for agricultural sciences students; medical diagnosis and treatment for medicine students; and organizational and financial management for business administration students. Test results are reported using a scale from 0 to 300, resulting from an IRT analysis of the data.

Analysis

Preparation of the data

One critical aspect to consider when designing SOMs is the scaling of the input variables (Kohonen, 2014). Any input variable whose scale is larger than the other scales can appear to be "more important" to the network, and consequently, produce misleading representations of the data. Unless a robust theoretical reason indicates that the importance of the input variables is different, they all should be considered as equally important. Therefore, in this study, all the input variables were normalized between 0 and 1 for further analysis. Values closer to 0 indicated a low

value of the variable (e.g., low-income, low prior achievement). Conversely, values closer to 1 represented a high value of the variable.

Designing the "best" SOM model

The software tool chosen for this study was the SOM Toolbox 2.0 under MATLAB 9.7 (R2019b). Although several software packages exist to create SOMs, Kohonen (2014) strongly recommends the use of the SOM Toolbox, developed by his research lab at the Helsinki University of Technology. The major advantage of using the SOM Toolbox is the flexibility of the MATLAB environment, as its functions make it possible to design SOMs in a simple and easy way (Vesanto, Himberg, Alhoniemi, & Parhankangas, 2000). Furthermore, SOMs were designed in four steps following the suggestions of Vesanto et al. (2000).

In step 1, a rectangular grid of 200 neurons (20 rows x 10 columns) was created. Such selection for the grid size is explained next. The SOM Toolbox calculates the default number of neurons of the grid using the expression 5*sqrt(n), where n is the number of cases in the data set (Kohonen, 2014). As there were 162,030 cases in the analyzed data set, the number of neurons for the grid would be approximately 2000. However, the design's computational complexity increases for a large number of neurons, which in practice could impact the training time as well as the interpretability of the resulting SOM. Therefore, the default number of neurons was reduced to 200.

Regarding the size of the grid, Kohonen (2014) has suggested that the relationship between the sides of a rectangular map (i.e., length and width) can be determined as the ratio between the two largest eigenvalues of the training data set. In the analyzed data set, this ratio was equal to

2.43. Consequently, the relationship between the length (l) and the width (w) of the map was represented using the two by two equations system shown in Equation 2.

$$\begin{cases} lw = 200 \\ l = 2.43w \end{cases} \quad (2)$$

Solving Equation 2 showed that the width of the grid should be 9.07, and the length of the grid should be 22.04. Both values were then respectively rounded to 10 and 20 so that the number of neurons in the grid was exactly 200. In step 2, the combination of two types of training (sequential and batch training), two ways of initialization (random and linear), and four types of neighborhood functions (Gaussian, cut gaussian, bubble, and Epanechnikov) yield the creation of 16 SOMs. In step 3, each of the created SOM was trained during 200 epochs. It is important to note that in the case of the sequential training, the initial learning rate was equal to 0.5, and then it decreased towards values close to 0. Conversely, no learning rate was needed in the case of batch learning (Kohohen, 2013). In step 4, the quality measures (quantization error and topographical error) were calculated for each of the trained SOMs. The SOM, which exhibited the best quality after considering all of the measures, was kept for further analysis.

Furthermore, the "hits map" of the best SOM model was created. A "hits-map" shows how many students are mapped to each of the neurons of the SOM (Kohonen, 2014). Therefore, the "hits-map" could be an indication of the distribution of the data set into the SOM (Kohonen, 2013). Although merely informative, the "hits-map" also corroborates whether the size of the SOM is adequate. More specifically, areas of the map with many students could suggest increasing the size of the grid, while areas of the map with few students could imply that a reduction of the grid is needed.

Patterns between academic performance, SES, and prior academic achievement

To answer the first research question posed in this study, both correlation analysis and visual analysis were carried out on the "best" SOM model. Regarding the correlation analysis, notice that the weight vectors in each one of the nodes correspond to the final values of the input variables after training the "best" SOM model. Therefore, the existing patterns among the input variables were established by calculating Pearson's r coefficient among the final weight vectors. The resulting correlation coefficients provided a quantitative indicator of the existing patterns between prior academic achievement, SES, and academic performance in higher education.

Regarding the visual analysis, the "component planes" for all the input variables were created. A "component plane" specifies individual information on the value of each input variable in every node of students on the map (Kohonen, 2014). In the "component plan", the magnitude of the variables is also represented by colors. Thus, darker colors indicate nodes where the variable is high, while lighter colors represent nodes where the variable is low. Interestingly, the comparison of the 12 resulting "component planes" provided a qualitative description of the existing patterns among the input variables.

Clustering students based on their prior academic achievement, SES, and academic performance

To answer the second research question of this study, both visual analysis and k-means clustering of the map units of the "best" SOM model was carried out. Concerning the visual analysis, the so-called "U-matrix" of the "best" SOM model was created. A "U-matrix" indicates how far or close the neurons are in the resulting SOM (Kohonen, 2014). In the "U-matrix", the distance between the neurons is represented by colors. The more similar the colors, the more

similar are the distances between the neurons. The “U-matrix” suggests the likely existence of clusters among the map units.

Regarding the k-means clustering, the procedure in the SOM Toolbox runs multiple times for each k, and the best run is chosen based on the sum of squares. Then, the optimal number of clusters is evaluated through the Davies-Bouldin index. The solution with the lowest index is selected as the best one. In this study, k was set to 10, and each k was run 1000 times. The Davies-Bouldin index was found to be the lowest for the solution of three clusters. Next, descriptive statistics of the three resulting clusters were performed. More specifically, the mean and standard deviation of each weight vector was calculated within each cluster. Subsequently, the mean for each input variable was compared among the three clusters through a Welch’s ANOVA. The Welch’s ANOVA is an alternative test to the classic one-way ANOVA, especially useful in case of heterogeneity of variance (Moder, 2010). Because there were significant differences between the three clusters means, a post hoc comparison using the Games-Howell test was also conducted. The Games-Howell test is recommended in cases where the assumption of homogeneity of variance is not fulfilled (Lee & Lee, 2018).

Results

The first part of this section describes the procedure followed to design the “best” SOM model. The second part of this section addresses the research question number one. The third part of this section answers the research question number two.

The design of the “best” SOM model

Table 1 summarizes the outcomes of the SOMs design process. Information regarding neighborhood function, initialization algorithm, type of training, and quality measures is provided.

According to Table 1, the smallest topographic errors were found for the "Gaussian" function, and the largest errors were found for the "Epanechnikov" function. This finding might suggest that the "Gaussian" function preserves better the relations among the input variables, particularly when they are scaled between 0 and 1. In addition, the quantization error was similar in each of the neighborhood functions. This finding seems to indicate that the quantization error depends more on several other parameters of the SOM, for instance, the grid size (Breard, 2017; Pölzlbauer, 2004).

Regarding the initialization algorithm (i.e., linear or random) and the type of training (i.e., sequential or batch), no remarkable differences were identified in the trained maps based on the calculated quality measures. In this respect, Akinduko, Mirkers, and Gorban (2016) have reported that there is no conclusive evidence on the better performance of linear initialization over random initialization. Similarly, Kohonen (2013) has indicated that the main difference between the two types of initialization is that through linear initialization the SOM converges faster to the final representation of the input data.

Concerning the type of training, Kohonen (2013) has also pointed out that batch training is a learning-rate free design, which shows a faster and safer convergence than sequential training. After training, the configuration which exhibited the least topographic error was chosen as the "best" SOM model, as this quality measure was found to be more sensitive to the changes of the parameters. Such configuration is the third one (map 3) indicated in Table 1: "Gaussian" function, random initialization, and sequential training.

Figure 2 shows the "hits-map" of the "best" SOM model. Notice that the "hits-map" indicates how many students were mapped to each of the neurons of the SOM. According to the "hits-map", there were no empty nodes in the "best" SOM model. Also, most of the neurons

located in the outermost part of the map are the most populated ones (more than 1000 students). In contrast, most of the neurons located towards the inside of the map are less populated (less than 100 students). This relatively uniform distribution of the students across the map may indicate that the chosen map size (20 rows x 10 columns) was adequate for representing the input data.

Which patterns exist between prior academic achievement, SES, and academic performance?

Table 2 shows the results of the correlation analysis carried out on the weight vectors across the 200 units of the "best" SOM model. According to Table 2, there is a very strong significant relationship among the scores in all the areas of the SABER 11 test and the SABER PRO test: Spanish ($r(200)= .995$, $p < .01$); mathematics ($r(200)= .989$, $p < .01$); social sciences ($r(200)= .988$, $p < .01$); philosophy ($r(200)= .981$, $p < .01$); biology ($r(200)= .992$, $p < .01$); chemistry ($r(200)= .987$, $p < .01$); and physics ($r(200)= .979$, $p < .01$). In addition, there is a moderate significant relationship among stratum ($r(200)= .576$, $p < .01$), income ($r(200)= .525$, $p < .01$), mother's educational level ($r(200)= .511$, $p < .01$), and father's educational level ($r(200)= .537$, $p < .01$) and the SABER PRO test. As expected, the relationship between income and stratum ($r(200)= .895$, $p < .01$) was strong and significant as well as the relationship between mother's educational and father's educational levels ($r(200)= .743$, $p < .01$).

Figure 3 presents the "component planes" for each one of the 12 input variables analyzed in this study. The "component plan" separately indicates the value of a variable in each one of the nodes forming the SOM. The analysis of the overlapping of the "component planes" reveals interesting patterns of the relationship among the input variables. The first pattern is observed in the upper left corner of the map. This area of the map groups the students who are the best performers in the SABER PRO test. These students also come from the highest stratum

houses, families with the highest income, their parents have the highest educational level, and have the highest performance in all the areas of the SABER 11 test.

The second pattern occurs in the lower right corner of the map. The students with the lowest performance in the SABER PRO test are placed in this area of the map. Besides, these students come from the lowest stratum houses, families with the lowest income, their parents have the lowest educational level, and have the lowest performance in all the areas of the SABER 11 test.

The third pattern appears in the center of the map. Remarkably, there are students with very high performance in the SABER PRO test, who also come from low to medium stratum houses, families with low income, have parents with low to medium educational level, and have very high performance in all the areas of the SABER 11 test. Altogether, these three first patterns corroborate the moderate association between the SES indicators (stratum, income, mother education, and father education) and the SABER PRO test found in the correlation analysis.

The fourth pattern is the very similar color distribution existing in all the areas of the SABER 11 test and the SABER PRO test. The better the result in the SABER 11 test, the better the result in the SABER PRO test. Therefore, this fourth pattern confirms the strong association found in the correlation analysis between both tests.

Summing up the results of both correlation and visual analysis, it can be stated that prior academic achievement and SES differ in their relation to academic performance in higher education. As such, the relationship between academic performance and prior academic achievement is more *linear*. Moreover, prior academic achievement is more strongly related to academic performance in higher education than SES.

Which clusters of students exist based on their prior academic achievement, SES, and academic performance?

Figures 4 (left side) displays the resulting "U-matrix" for the "best" SOM model. The "U-matrix" shows in colors the distances among the 200 neurons forming the grid. A visual inspection of the "U-matrix" reveals that there are three areas of the map where the distances are dissimilar, which might indicate the existence of at least three clusters of nodes. In Figure 4, the three presumed areas are shown in yellow, green, and blue. To further confirm the existence of three clusters of students, k-means clustering was performed on the 200 units of the "best" SOM model. After 1000 runs, the David-Bouldin index was calculated for each of the chosen values of k. The smaller the index, the better the grouping. According to this index, the optimal number of groups of students is three, as displayed in Figure 4 (right side).

Figure 5 (left side) shows in colors, each one of the three clusters of students and bars (right side) the number of students classified within each cluster. The yellow cluster was labeled as "high"; the blue cluster was labeled as "medium", and the red cluster was labeled as "low", as it is explained subsequently. To begin with, the mean and standard deviation were calculated for each of the 12 input variables in the three clusters of students as shown in Table 3. According to Table 3, the "high" group includes the students who, on average, have a high performance in the SABER PRO test, high performance in all the areas of the SABER 11 test, and high values in the SES indicators (stratum, income, mother's educational and father's educational levels). In addition, the group "medium" comprises the students who, on average, exhibit a medium performance in the SABER PRO test, a medium performance in all the areas of the SABER 11 test, and have medium values in the SES indicators. Finally, the students belonging to the "low"

group are the students who, on average, have low performance in the SABER PRO test, low performance in all the areas of the SABER 11 test, and have low values in the SES indicators.

Next, a Welch's ANOVA was conducted to compare the means of the 12 input variables among the "high", "medium", and "low" clusters. The results from Welch's ANOVA showed that there were significant differences for all the means across the three clusters, as presented in Table 4. Because of these significant differences, a post hoc comparison using the Games-Howell test was conducted as reported in Table 5. The Games-Howell test showed that the mean value of each input variable in the "high" group was significantly higher than in the "medium" group and in the "low group". Similarly, the mean value of each input variable in the "medium" group was significantly higher than in the "low" group.

A concluding remark of these analyses is that three different groups of students exist, namely "high", "medium", and "low", based on their prior academic achievement, SES conditions, and academic performance in higher education. Besides, these variables are statistically different among the three existing groups.

Discussion

The objective of the present study was twofold. The first objective was to analyze the existing patterns between prior academic achievement, SES, and academic performance in a cohort of 162,030 Colombian university students through a SOM model. The second objective was to cluster the resulting SOM to determine specific groups of students based on their prior academic achievement, SES, and academic performance in higher education.

The present study contributes to the existing literature on higher education in three specific ways. First, this study shows that it is possible to analyze large educational data sets using artificial

neural network models and to obtain meaningful interpretations out of it. As such, this study illustrates how unsupervised learning models can discover patterns of information, even when no "input-output" relationship is specified. This fact invites educational researchers to consider several other possibilities to do quantitative research, which are complementary to the classical statistical techniques. Second, this study suggests that through exploratory analyses such as SOMs, a better understanding of academic performance in higher education could be achieved. In particular, the proposed SOM in this study led to finding interesting patterns among prior academic achievement, SES, and academic performance in higher education. Third, this study encourages educational researchers to explore alternative and more robust methods for comparing means across groups such as Welch's ANOVA and the Games-Howell test. This fact is particularly important given that violating the assumption of equal variance across the groups causes Type I error to inflate, resulting in misleading conclusions.

Regarding the first research question, the "best" SOM model revealed insights about the existing patterns between prior academic achievement, SES, and academic performance in higher education. To begin with, the findings of this study indicated that a linear pattern exists between prior academic achievement and academic performance in higher education. Although this finding was somehow expected, it also suggests the importance of commencing university studies with adequate academic preparation. In addition, there was evidence of a non-linear association between SES and academic performance in higher education. This finding refutes the idea that only students coming from most privileged families would perform well in the university. As a matter of fact, there were students placed on the center of the "best" SOM model who achieved very high performance in the SABER PRO test and come from the least advantaged families (low stratum, low income, low parents' educational level). Interestingly, such students also exhibited

very high performance in all the areas of the SABER 11 test. As this study confirms that prior academic achievement is stronger related to academic performance in higher education than SES, a first step to support low-SES students would be to offer them preparation programs to foster their academic preparation before university.

Regarding the second research question, three clusters of students, namely, “high”, “medium”, and “low” exist in the resulting SOM model. Each cluster grouped the students in relation to their prior academic achievement, SES, and academic performance in higher education. The means of these input variables were statistically different among the three clusters, which may indicate their usefulness to distinguish university students efficiently. More considerable differences across the clusters were found in the SES indicators (stratum, income, parents’ educational level), whereas the smallest differences were observed in the outcomes of both SABER 11 and SABER PRO test. This fact is consistent with the larger standard deviation observed in the SES indicators. Therefore, findings of this study seem to suggest that both SES and prior academic achievement could be considered as predictors of academic performance in higher education.

Finally, some hints for designing SOMs can be proposed after carrying out this study. The first one is to normalize the input variables so that they have the same importance to the model. The use of the original scaling of the input variables may lead to the identification of erroneous patterns in the data. The second one is to train several SOM models by systematically change their tuning parameters. In this way, more valid representations of the input space could be found. The third one is to calculate several quality measures and not rely exclusively on the quantization error. In this respect, the topographic error could be a more interesting value for educational researchers,

as it represents how well the existing relationship among the input variables is preserved in the final model.

Limitations

Several limitations of this study must be acknowledged. The first limitation is that the operationalization of both prior academic achievement and academic performance in higher education was based on standardized test outcomes only. Given the size of the analyzed data set, information of more classical indicators (e.g., grade point average) could not be retrieved for the analysis. The second limitation is that only quantitative variables were selected for the analysis. Although there are fascinating examples of the use of categorical variables in SOMs (e.g., Hsu, 2006; Hsu & Lin, 2011), the design of such models was beyond the scope of this study. Therefore, meaningful categorical variables such as parents' occupation or students' academic programs were not included in the analysis. The third limitation is that some of the identified patterns in the SOM model could only occur in the Colombian educational setting; thus, they could be less generalizable.

Future research

The use of SOMs could shape future research in higher education in several ways. As a clustering analysis tool, SOMs could be used to discover patterns of teaching and learning existing at the individual, classroom, and institutional levels. This fact would be of enormous benefit, as it would suggest what works best to foster students' learning process. Furthermore, SOMs could be used to understand the academic and behavioral patterns of students who drop out of university. By following these patterns, several strategies to reduce academic desertion from the university could be proposed. Finally, the use of SOMs could extend the development of *real* predictive

models of academic performance in higher education, which do not require the specification of any “input-output” relationship. In this respect, the use of the so-called super SOMs (i.e., SOMs with more than one competitive layer) should be explored.

References

Akinduko, A. A., Mirkes, E. M., & Gorban, A. N. (2016). SOM: Stochastic initialization versus principal components. *Information Sciences, 364,* 213-221. doi:10.1016/j.ins.2015.10.013

Alias, U. F., Ahmad, N. B., & Hasan, S. (2015). Student behavior analysis using self-organizing map clustering technique. *ARPN Journal of Engineering and Applied Sciences, 10,* 17987-17995.

Antonenko, P. D., Toy, S., & Niederhauser, D. S. (2012). Using cluster analysis for data mining in educational technology research. *Educational Technology Research and Development, 60,* 383-398. doi: 10.1007/s11423-012-9235-8

Baker, R. S. J. D. (2010). Data mining for education. In McGaw, B., Peterson, P., Baker, E. (Eds.) *International Encyclopedia of Education*, Vol.7 (pp.112-118). Oxford, UK: Elsevier.

Breard, Gregory T. (2017). *Evaluating self-organizing map quality measures as convergence criteria* (Master's thesis). Retrieved from: http://digitalcommons.uri.edu/theses. (Paper No. 1033).

Daniel, B. K. (2017). *Big data and learning analytics in higher education*. Switzerland: Springer International Publishing, doi: 10.1007/978-3-319-06520-5

Da Silva, E. T., de Fátima Nunes, M., Santos, L. B., Queiroz, M. G., & Leles, C. R. (2012). Identifying student profiles and their impact on academic performance in a Brazilian undergraduate student sample. *European Journal of Dental Education, 16,* e27-e32. doi: 10.1111/j.1600-0579.2010.00669.x

De Clercq, M., Galand, B., & Frenay, M. (2017). Transition from high school to university: A person-centered approach to academic achievement. *European Journal of Psychology of Education, 32,* 39-59. doi:10.1007/s10212-016-0298-5

Dutt, A., Aghabozrgi, S., Ismail, M. A. B., & Mahroeian, H. (2015). Clustering algorithms applied in educational data mining. *International Journal of Information and Electronics* Engineering, *5*, 112-116. doi: 10.7763/ijiee.2015.v5.513

Garson, G. D. (2014). *Neural network models*. Asheboro, NC: Statistical Associates Publishers.

Gibson, D., & Ifenthaler, D. (2017). Preparing the next generation of education researchers for big data in higher education. In Daniel B.K. (Eds), *Big data and learning analytics in higher education* (pp. 29-43). Switzerland: Springer International Publishing, doi: 10.1007/978-3-319-06520-5

Haykin, S. S. (2009). *Neural networks and learning machines* (Vol. 3). Upper Saddle River, NJ, USA: Pearson.

Hsu, C. C. (2006). Generalizing self-organizing map for categorical data. *IEEE Transactions on Neural Networks, 17,* 294-304. doi: 10.1109/tnn.2005.863415

Hsu, C. C., & Lin, S. H. (2011). Visualized analysis of mixed numeric and categorical data via extended self-organizing map. *IEEE Transactions on Neural Networks and Learning Systems, 23,* 72-86. doi: 10.1109/tnnls.2011.2178323

Huberty, C. J., Jordan, E. M., & Brandt, W. C. (2005). Cluster analysis in higher education research. In Smart J.C. (Eds), *Higher education: Handbook of Theory and Research: Vol. 20* (pp. 437-457). Springer: Dordrecht. doi: 10.1007/1-4020-3279-x_8

Iturriaga, F. J. L., & Sanz, I. P. (2013). Self-organizing maps as a tool to compare financial macroeconomic imbalances: The European, Spanish and German case. *The Spanish Review of Financial Economics, 11*, 69-84. doi: 10.1016/j.srfe.2013.07.001

Kausar, S., Huahu, X., Hussain, I., Wenhao, Z., & Zahid, M. (2018). Integration of data mining clustering approach in the personalized E-learning system. *IEEE Access, 6*, 72724-72734. doi: 10.1109/access.2018.2882240

Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. *Biological Cybernetics,43,* 59-69. doi: 10.1007/BF00337288

Kohonen, T. (2013). Essentials of the self-organizing map. *Neural Networks, 37*, 52-65. doi: 10.1016/j.neunet.2012.09.018

Kohonen, T. (2014). *MATLAB implementations and applications of the self-organizing map*. Helsinki, Finland: Unigrafia Oy.

Kumar, U. A., & Dhamija, Y. (2010). Comparative analysis of SOM neural network with K-means clustering algorithm. *Proceedings of the IEEE International Conference on Management of Innovation & Technology 2010,* 55-59. doi: 10.1109/icmit.2010.5492838

Lee, S., & Lee, D. K. (2018). What is the proper way to apply the multiple comparison test? *Korean Journal of Anesthesiology, 71,* 353-360. doi: 10.4097/kja.d.18.00242

Lee, Y. (2019). Using self-organizing map and clustering to investigate problem-solving patterns in the massive open online course: an exploratory study. *Journal of Educational Computing Research, 57,* 471-490. doi: 10.1177/0735633117753364

Moder, K. (2010). Alternatives to F-test in one-way ANOVA in case of heterogeneity of variances (a simulation study). *Psychological Test and Assessment Modeling, 52,* 343-353.

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*, e99478. doi: 10.1371/journal.pone.0099478

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*, 436-448. doi: 10.1093/her/cyv016

Pölzlbauer, G. (2004). Survey and comparison of quality measures for self-Organizing Maps. In J. Paralič, G. Pölzlbauer & A. Rauber (Eds.), *Proceedings of the Fifth Workshop on Data Analysis (WDA'04)* (pp. 67-82). Vysoké Tatry, Slovakia: Elfa Academic Press.

Qi, J., Yu, Y., Wang, L., Liu, J., & Wang, Y. (2017). An effective and efficient hierarchical K-means clustering algorithm. *International Journal of Distributed Sensor Networks*, *13*, 1-17. doi: 10.1177/1550147717728627

Rodríguez, C. F. (2016). *Socioeconomic factors and outcomes in higher education: A multivariate analysis*. Bogotá: Universidad Externado de Colombia. doi:10.4000/books.uec.1496

Romero, C., & Ventura, S. (2010). Educational data mining: review of the state of the art. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 40,* 601-618. doi: 10.1109/TSMCC.2010.2053532

Shahiri, A. M., Husain, W., & Rashid, N.A. (2015). A review on predicting student's performance using data mining techniques. *Procedia Computer Science, 72,* 414-422, doi: 10.1016/j.procs.2015.12.157

Sirin, S. (2005). Socioeconomic status and academic achievement: A meta-analytic review of research. *Review of Educational Research, 75,* 417-453. doi:10.3102/00346543075003417

The World Bank (2012). *Reviews of national policies for education: Tertiary education in Colombia 2012*. Paris: OECD Publishing. doi:10.1787/9789264180697-en

Van Ewijk, R., & Sleegers, P. (2010). The effect of peer socioeconomic status on student achievement: A meta-analysis. *Educational Research Review, 5,* 134-150. doi:10.1016/j.edurev.2010.02.001

Vesanto, J., Himberg, J., Alhoniemi, E., & Parhankangas, J. (2000). *SOM Toolbox for Matlab 5* (Research Report No. A57). Retrieved from Helsinki University of Technology, Laboratory of Computer and Information Science website: http://www.cis.hut.fi/projects/somtoolbox/documentation/

Wehrens, R., & Buydens, L. M. (2007). Self-and super-organizing maps in R: the Kohonen package. *Journal of Statistical Software, 21,* 1-19. doi: 10.18637/jss.v021.i05

Table 1. Training of SOM models

Map number	Neighborhood function	Initialization	Training algorithm	Quantization error	Topographic error
1	Gaussian	Linear	Sequential	.25	.06
2			Batch	.25	.05
3		Random	Sequential	.26	.03
4			Batch	.26	.04
5	Cutgauss	Linear	Sequential	.24	.09
6			Batch	.24	.11
7		Random	Sequential	.24	.09
8			Batch	.24	.08
9	Bubble	Linear	Sequential	.24	.07
10			Batch	.24	.10
11		Random	Sequential	.24	.08
12			Batch	.24	.07
13	Epanechnikov	Linear	Sequential	.22	.39
14			Batch	.22	.35
15		Random	Sequential	.23	.11
16			Batch	.23	.13

Table 2. Descriptive statistics and correlation of the weight vectors (n= 200)

Variable	***M***	***SD***	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**
1. Stratum	.46	.13											
2. Income	.32	.20	.895*										
3. Mother's education	.6	.20	.671*	.621*									
4. Father's education	.58	.24	.6*	.582*	.743*								
5. S11 Spanish	.55	.04	.563*	.514*	.483*	.5*							
6. S11 mathematics	.42	.04	.593*	.559*	.521*	.53*	.991*						
7. S11 social sciences	.49	.04	.501*	.453*	.431*	.451*	.992*	.985*					
8. S11 philosophy	.46	.04	.459*	.409*	.401*	.442*	.988*	.978*	.996*				
9. S11 biology	.42	.03	.522*	.484*	.476*	.48*	.995*	.991*	.994*	.991*			
10. S11 chemistry	.43	.03	.535*	.497*	.469*	.484*	.993*	.993*	.991*	.988*	.996*		
11. S11 physics	.39	.03	.518*	.491*	.469*	.48*	.986*	.991*	.985*	.982*	,991*	.997*	
12. SABER PRO	.52	.05	.576*	.525*	.511*	.537*	.995*	.989*	.988*	.981*	.992*	.987*	.979*

Note. *M* and *SD* are used to represent mean and standard deviation, respectively. * indicates p <.01.

Table 3. Descriptive statistics for each cluster

Cluster		Stratum	Income	Mother's education	Father's education	S11 Spanish	S11 mathematics	S11 social sciences	S11 philosophy	S11 biology	S11 chemistry	S11 physics	SABER PRO
High (n=34)	*M*	.68	.66	.84	.86	.59	.46	.52	.5	.45	.46	.42	.56
	SD	.06	.17	.09	.1	.03	.04	.03	.03	.03	.03	.03	.03
Medium (n=75)	*M*	.47	.32	.69	.71	.56	.43	.5	.48	.43	.44	.4	.54
	SD	.08	.11	.15	.16	.03	.03	.03	.03	.02	.02	.02	.03
Low (n=91)	*M*	.37	.19	.44	.36	.52	.39	.47	.44	.4	.41	.38	.49
	SD	.08	.06	.12	.12	.04	.04	.04	.04	.03	.03	.03	.04

Note. M and *SD* are used to represent mean and standard deviation, respectively.

Table 4. Welch's ANOVA for the clusters means

Variable	*Welch*	*df1*	*df2*	*p*
Stratum	244.535	2	100.783	.000
Income	161.867	2	71.046	.000
Mother's education	212.470	2	106.424	.000
Father's education	320.810	2	97.908	.000
S11 Spanish	49.926	2	93.328	.000
S11 mathematics	51.661	2	85.694	.000
S11 social sciences	37.351	2	92.242	.000
S11 philosophy	32.575	2	91.313	.000
S11 biology	44.919	2	91.093	.000
S11 chemistry	45.967	2	87.841	.000
S11 physics	42.227	2	85.355	.000
SABER PRO	55.300	2	94.065	.000

Table 5. Games-Howell test for multiple comparison across the clusters means

Variable	(I) Cluster	(J) Cluster	`Mean difference (I-J)	S.E.
Stratum	High	Low	0.31**	0.01
		Medium	0.21**	0.01
	Medium	Low	0.10**	0.01
Income	High	Low	0.47**	0.03
		Medium	0.34**	0.03
	Medium	Low	0.13**	0.01
Mother's education	High	Low	0.40**	0.02
		Medium	0.15**	0.02
	Medium	Low	0.25**	0.02
Father's education	High	Low	0.50**	0.02
		Medium	0.15**	0.03
	Medium	Low	0.35**	0.02
S11 Spanish	High	Low	0.07**	0.01
		Medium	0.03**	0.01
	Medium	Low	0.04**	0.01
S11 mathematics	High	Low	0.07**	0.01
		Medium	0.03**	0.01
	Medium	Low	0.04**	0.01
S11 social sciences	High	Low	0.05**	0.01
		Medium	0.02**	0.01
	Medium	Low	0.03**	0.01
S11 philosophy	High	Low	0.06**	0.01
		Medium	0.02*	0.01
	Medium	Low	0.04**	0.01
S11 biology	High	Low	0.05**	0.01
		Medium	0.02**	0.01
	Medium	Low	0.03**	0.00
S11 chemistry	High	Low	0.05**	0.01
		Medium	0.02**	0.01
	Medium	Low	0.03**	0.00
S11 physics	High	Low	0.04**	0.01
		Medium	0.02**	0.01
	Medium	Low	0.02**	0.00
SABER PRO	High	Low	0.07**	0.01
		Medium	0.02**	0.01
	Medium	Low	0.05**	0.01

Note. ** indicates p <0.01; * indicates p <0.05

Figure 1. Architecture of a self-organizing map (SOM)

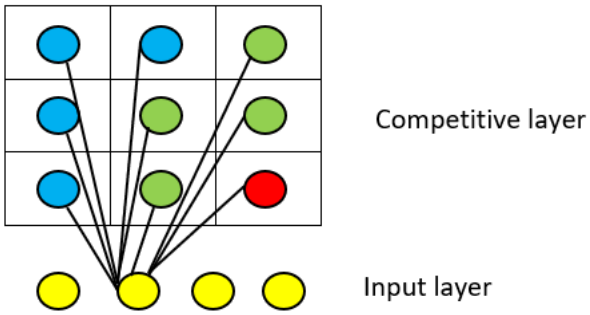


Figure 2. “Hits map” for the “best” SOM model

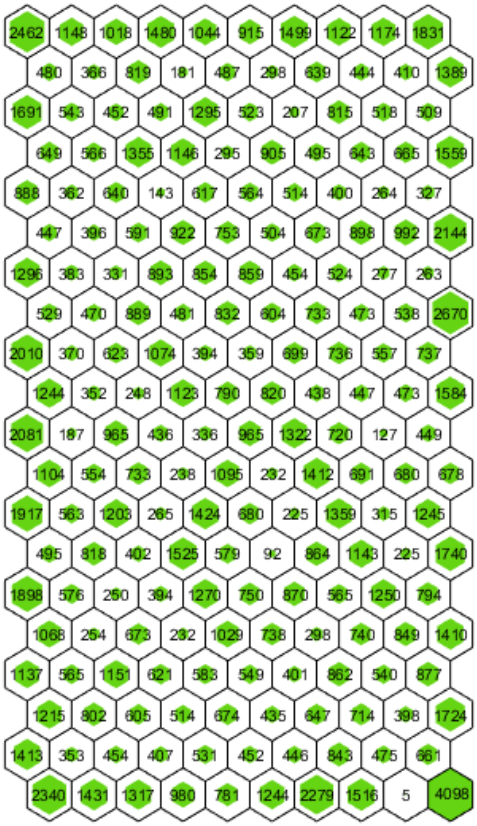


Figure 3. “Component planes” for the input variables

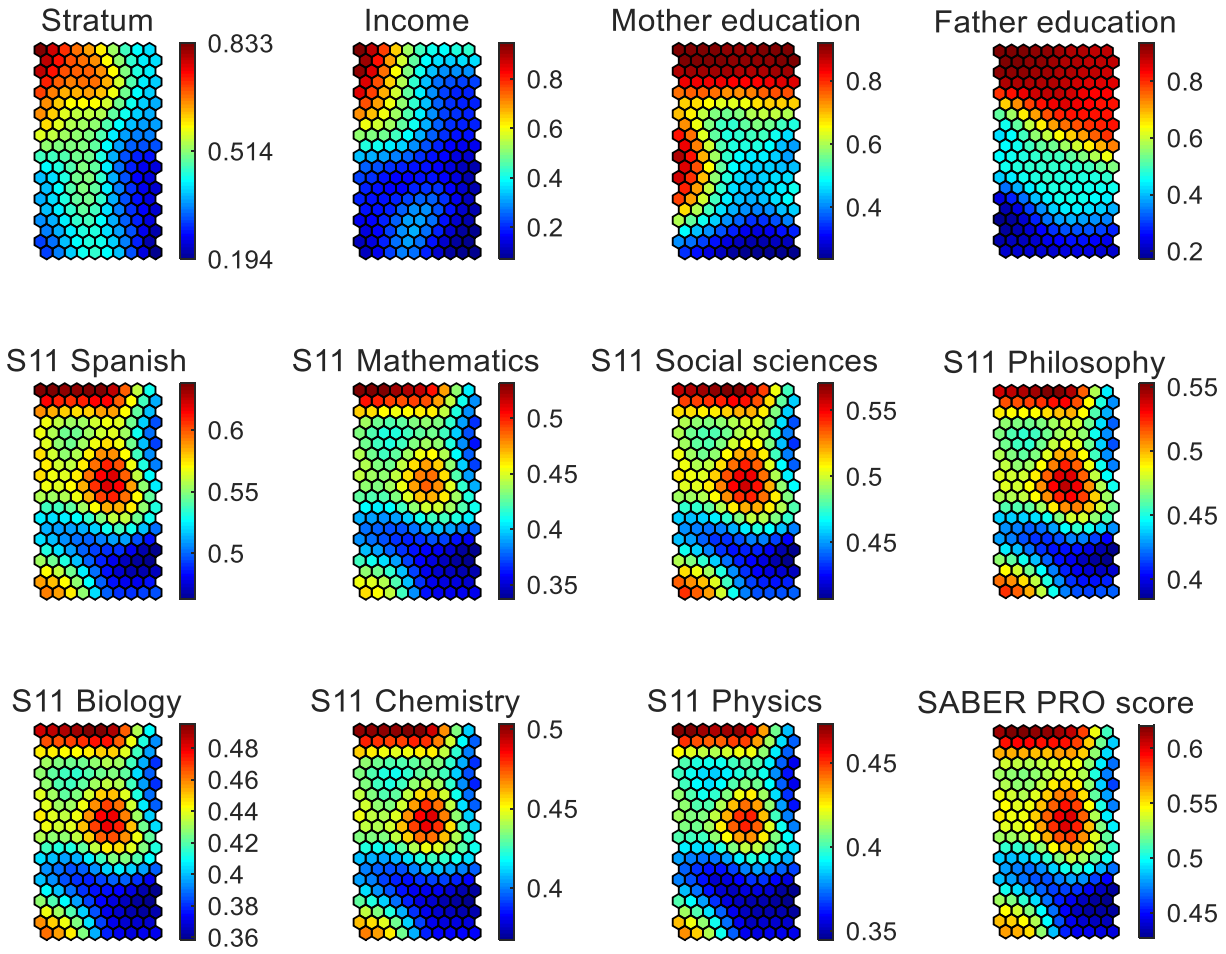


Figure 4. “U matrix” and selection of the optimal number of clusters

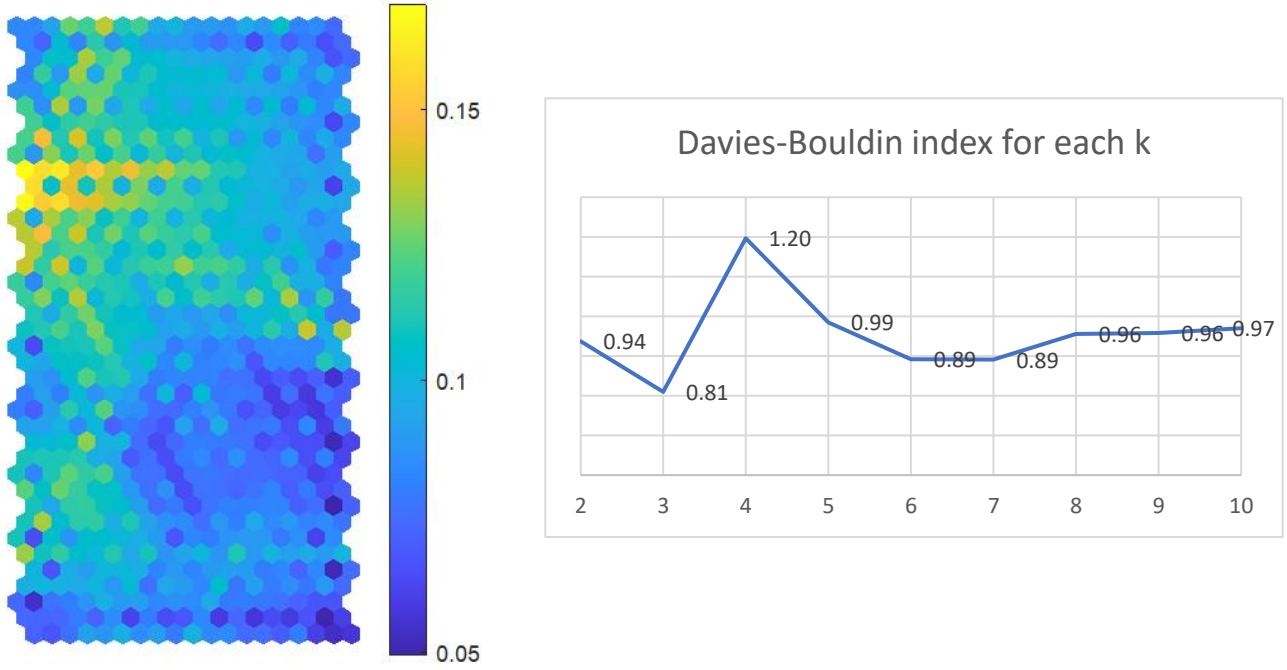
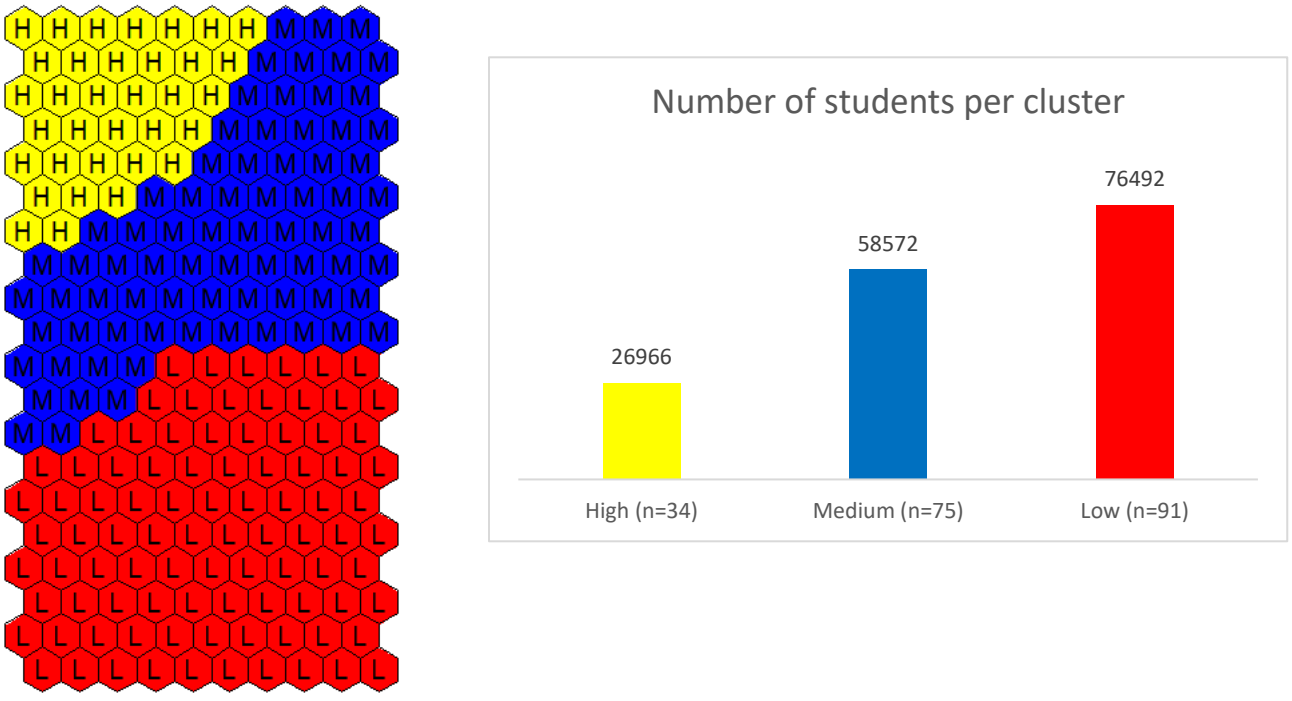


Figure 5. Clusters of students based on the input variables



Chapter 6

General Discussion

"The original question, 'Can machines think?' I believe to be too meaningless to deserve discussion." (Alan Turing)

The analysis of academic performance is not just another phenomenon for educational scientists to research. To achieve a better understanding of the factors which play a role in determining students' academic performance is of benefit for students, teachers, institutions, and policymakers. Up until now, educational researchers have mainly relied on the use of traditional statistical techniques when analyzing educational data. Nevertheless, the increase of the available information regarding students' performance as well as the development of more sophisticated computational tools invite educational researchers to explore new ways of analysis, as these sophisticated computational tools enable a fast and precise analysis of vast volumes of data, highly diverse in their type, structure, and level of measurement. In this way, the complex interactions existing within those vast volumes of data can be identified and interpreted.

Therefore, the analysis of academic performance should include not only elements inherent to educational sciences but also incorporate elements coming from other disciplines. In other words, educational researchers need to adopt an interdisciplinary approach when investigating academic performance in higher education. To date, educational Data Mining (EDM) is possibly the best example of such an interdisciplinary approach. Within the existing data mining techniques used in EDM, artificial neural networks (ANNs) emerge as a powerful technique to properly analyze educational data. Remarkably, ANNs help to develop robust predictive systems

independent of the level of measurement and the statistical distribution of the data, even when the investigated variables are highly correlated.

This doctoral dissertation aimed to extend both conceptually and methodologically the understanding of ANN, so their use can be increased among educational researchers. As an illustrative case of the use of ANNs to analyze academic performance in higher education, a unique data set coming from the Colombian educational setting and consisting of 162,030 students was investigated. As a first step, a mixed-methods systematic literature review was conducted to further extend the understanding of the existing literature on academic performance in higher education (study 1). Subsequently, through both supervised and unsupervised ANNs architectures, three major tasks were performed in relation to the analysis of academic performance in higher education: classification (study 2), regression (study 3), and clustering (study 4). In the following sections, the key findings of this doctoral dissertation are listed and discussed. Next, the conclusions and implications for practice, as well as the limitations of this research findings, are highlighted. Finally, suggestions for future research are given.

Key findings of this doctoral dissertation

Key finding 1: The relationship between SES and academic performance in higher education is positive yet weak.

The meta-analysis conducted in study 1 revealed a small average effect size for the SES-academic performance relationship. Furthermore, findings from studies 2, 3, and 4, also revealed that SES was less strongly related to academic performance in higher education than the other predictors. Nevertheless, these findings do not suggest ignoring the influence of SES on academic performance in higher education, as there are several reasons which could explain the weakness

found in this relationship. Firstly, the influence of SES on academic performance seems to be larger in levels of education preceding university (e.g., Sirin, 2005; Van Ewijk & Sleegers, 2010). Secondly, it can be argued that once low-SES students are admitted to the university and commence their studies, they experience the new educational setting in a similar way as their high-SES counterparts (Smith, 2016).

The study of the relationship between SES and academic performance also requires considering how these constructs are measured. In this respect, study 1 proposed conceptual models for measuring both SES and academic performance. SES should be operationalized using five measurements: education, occupation, income, home resources, and neighborhood resources. Moreover, such measures exist in three possible levels, namely, the individual, family or area levels. To dismiss the multidimensional nature of SES as well as its different levels of measurement would lead to its incorrect use and interpretation as a predictor of academic performance. In particular, the selection of SES indicators at the area level to describe students' socio-economic status is problematic, as it leads to overestimating the predictive role of SES.

Furthermore, academic performance in higher education should be assessed in terms of achievement, competencies, and persistence, both at university and before university. Therefore, it seems adequate to explore several other operationalizations of academic performance different from grade point average (GPA). More specifically, valuable information regarding students' academic performance is defined by what they can do with the acquired knowledge (competencies) and their progression towards the completion of their degrees (persistence). Also, the proposed conceptual model indicates the need to recognize the so-called temporal dimension when measuring academic performance in higher education. Consequently, more precise tracking of

students' academic progression could be achieved after contrasting the same measurements before and during university.

Key finding 2: Prior academic achievement, university experience, and working status are potential mediators of the relationship between SES and academic performance in higher education.

Study 1 also suggested the need for exploring how additional institutional and individual variables influence the relationship between SES and academic performance in higher education. Such finding is consistent with the I-E-O model (Astin, 1993), which indicates that SES is, directly and indirectly, related to academic performance in higher education. Interestingly, the presumed mediating role of the investigated variables points out specific ways to attend the educational needs of low-SES students. The mediating role of prior academic achievement suggests that preparation programs for university could help to manage the disadvantages of low-SES students. An example of such preparation programs is the enabling programs proposed by the Australian government, which provided specific competencies (Pitman et al., 2016) as well as effective pathways to university for low-SES students (Hodges et al., 2013).

Furthermore, the mediating role of university experience indicates that low-SES students can benefit not only from formal academic aspects such as teaching, learning environment, and assessment but also from extracurricular activities, social interaction with their peers, and participation in students' organizations. Therefore, universities should focus on how to integrate low-SES students into the new educational setting effectively. Finally, working during university could also positively influence low-SES students if the students have a part-time job related to their academic field (Wang, Kong, Shan, & Vong, 2010).

Key finding 3: The predictors of academic performance in higher education contribute differently to classify it as either high or low.

Study 2 made use of ANNs to reveal that prior academic achievement, SES, and high school characteristics do contribute to the classification of academic performance in higher education. Although the predictive contribution of working status was low, this fact is easily explained if one considers that the only included attribute describing working status was the number of hours worked. Study 2 also showed that university background (i.e., number of semesters completed, type of university, and academic program) had a considerable predictive role in the classification of academic performance. As the university background captures part of the students' university experience, there is an agreement with findings from study 1, which also identified the predictive role of the university experience.

Interestingly, the investigated predictors exhibited different prediction patterns in each performance group. For instance, the predictive contribution of prior academic achievement was larger in the low-performance group than in the high-performance group. This fact suggests that lower levels of prior academic achievement would result in lower levels of academic performance in higher education, regardless of the rest of the predictors. In addition, the predictive contribution of SES indicators was greater in the high-performance group than in the low-performance group. Thus, it can be argued that prior academic achievement shows less proportion of predictive power to classify academic performance as high, as other predictors also contribute to such a classification. Therefore, it can be concluded that the classification patterns did vary across the academic performance groups, given that more predictors with specific patterns are involved in the accurate classification of academic performance as high-performance.

Key finding 4: Predicting academic performance in higher education is not only about accuracy but also interpretability.

Study 3 responds to one of the major criticisms of the use of predictive models based on ANNs. More specifically, it has been said that although the resulting prediction coming from ANNs is highly accurate, how to correctly interpret it remains unclear. Thus, the general practice among educational researchers has been to use general linear methods to develop predictive models of academic performance. Nevertheless, it could be the case that such a practice is inconvenient when the investigated dataset includes thousands of cases and several nominal and ordinal variables. In this doctoral dissertation, a multiple linear regression would have led to a predictive model that includes 100 regression coefficients (analysis not shown here). However, the interpretation of 100 regression coefficients is not straightforward. Moreover, regression coefficients coming from dummy coded variables indicate the direction of the influence but not its effect size.

In contrast, study 3 showed that predictive systems based on ANNs do have both accuracy and interpretability, two defining characteristics of any successful predictive model. Notably, the predictive model developed in study 3 corroborated the ability of ANNs to approximate any input-output function assumed to be non-linear as the relationship found between academic performance and its predictors. These features become more relevant when analyzing large data sets containing several categorical predictors. Findings from Study 3 also indicated that prior academic achievement and SES were related to academic performance in higher education, despite the previous influence of SES on academic achievement before university.

Key finding 5: The relationship between academic performance in higher education and its predictors can be clustered in statistically different groups.

Study 4 illustrated how unsupervised ANNs could be utilized to identify existing patterns between academic performance in higher education and its predictors. Besides, study 4 showed that ANNs could reduce the high dimensionality spanned by several variables into a two-dimensional graphical representation where the association patterns become more evident. In this respect, findings from study 4 indicated that prior academic achievement was more strongly related to academic performance in higher education than SES. There was also evidence of a non-linear relationship between SES and academic performance, as expressed by the moderate correlation coefficient found between these two variables. This finding corroborates previous results of this doctoral dissertation, being possible to infer that not only students within the most privileged socio-economic settings perform well at university.

Also, study 4 demonstrated that it is possible to use k-means clustering to group the ANN representation of the existing patterns between academic performance in higher education and its predictors. As such, the identified groups were named following the level of each of the investigated variables within each group. Hence, the chosen labels were high, medium, and low. This ability of ANNs to significantly differentiate students' groups is of importance for the analysis of academic performance in higher education, given that it could help to establish the predictive role of a specified set of predictors.

Key finding 6: Building predictive models based on artificial neural networks is not a trial and error task.

Studies 2, 3, and 4 showed that it is possible to follow a systematic approach when designing ANNs, refuting the common understanding that ANNs mysteriously identify complex patterns in the analyzed data set. Such a systematic approach aimed to effectively determine the combination of hyper-parameters (e.g., learning rate, momentum, number of neurons, etc.), which maximizes the accurate classification, prediction, and clustering of academic performance in higher education. Regardless of the chosen topology, a simple followed step within this systematic design was to standardize the continuous variables through standard procedures (e.g., z-scoring, normalization), so the ANNs can equivalently analyze the variables. Although it is not possible to ensure that the "best" model would be identified through this systematic approach, its use would lead to reporting more generalizable results, as the same steps could be followed by any designer interested in similar predictive tasks. Therefore, the exploration of heuristic principles for the design of ANNs could be of benefit for the educational research, to achieve more accurate and valid predictive models.

Key finding 7: The assessment of the quality of any ANN model implies the calculation of several quality measures.

The three empirical studies conducted in this doctoral dissertation indicated the need for calculating various measures for assessing the quality of the ANNs models. To begin with, study 2 showed that less training error of the multilayer perceptron (MLP) does not necessarily imply the best performance during their testing. A low training error could be an indication of overfitting of the network, which would also lead to a poor generalization of the resulting model. In this respect, quality measures based on the confusion matrix of the classification (e.g., sensitivity,

specificity, F1 score) would allow more reliable identification of the hyper-parameters that maximize the desired classification. In the same way, study 3 indicated that even though the accuracy of the predictions of the RBFNNs models is calculated as the correlation coefficient between the expected value and the predicted value, one additional informative measure of the accuracy of the RBFNNs is the average overall relative error (relative to the mean model). Finally, study 4 showed that at least two types of error should be calculated when designing unsupervised artificial neural networks such as self-organizing maps (SOMs). In this case, it was a proper choice to calculate several quality measures and not rely exclusively on the quantization error. As such, the topographic error was of interest for this doctoral research, as it represented how well the existing relationship among the input variables was preserved in the final SOM model. In general, it can be argued that through the calculation of several quality measures, the validity of the predictive models based on ANNs could be increased, as these measures are consistent and complementary among them.

Conclusions and implications for research

The first conclusion is that the analysis of academic performance in higher education requires a non-linear approach. This conclusion is of enormous relevance, given that the general trend in the educational literature has been to develop predictive models of academic performance based on general linear models. As such, predictive models based on general linear methods narrow the way of relating the investigated variables. Alternatively, in predictive models based on ANNs, the predictors of academic performance in higher education can interact in several complex ways, apart from the linear one, to classify, predict, and cluster large educational data sets.

The second conclusion is that the analysis of academic performance in higher education should combine several research methods. The systematic literature review included in this

doctoral dissertation was conducted through a mixed approach (qualitative and quantitative). Such an approach allowed the valid combination of the results reported in the literature regarding academic performance in higher education. Furthermore, the quantitative approach underneath the ANNs made it possible to analyze a considerable amount of information, to include several predictors of academic performance, and to identify meaningful relations among them.

The third conclusion is that the use of artificial neural networks could be framed within a general science approach named model-dependent realism (Hawking & Mlodinow, 2010). According to this framework, a plausible model for analyzing academic performance in higher education would be the one that integrates several complementary partial models that provide information about the investigated phenomena, without being contradictory among them. As it was shown in this doctoral dissertation, the designing process of a predictive model based on ANNs implies the development of several partial models, derived from the systematic changes of specific parameters of the network (e.g., topology, activation function, learning paradigm).

The fourth conclusion is that previous criticisms that have prevented the use of ANNs in education research should be debated. To begin with, it has been criticized that no precise information about the input-output relationship exists in ANNs. However, this doctoral dissertation has shown how to use well-established algorithms to train three different topologies of ANNs. Thus, the input-output relationship was fully represented in the interactions among the neurons, which determine the predictive weights of the independent variables. Although this relationship was complex it existed in each of the reported models. Furthermore, it was possible to assess the quality of the classification, regression, and clustering based on ANNs, through specific quality measures. Finally, it was also shown that the predictive contribution of a given set

of variables could be determined and interpreted when ANNs are used to analyze academic performance in higher education.

The fifth conclusion is that ANNs truly consider the categorical nature of both nominal and ordinal predictors of academic performance in higher education. Unlike other predictive techniques, there is no numerical meaning in the dummy codification presented to the ANNs. As a matter of fact, in supervised topologies such as MLP or RBFNNs, each dummy variable is simply an input node within the input layer of the network. This suggests that ANNs possess the advantage of analyzing several important predictors of academic performance in higher education, usually measured using categorical scales.

Limitations

Although this doctoral dissertation has contributed to the field of educational sciences in many ways, three of its limitations must be acknowledged. The first limitation is that the analyzed database did not contain certain weighty predictors of academic performance in higher education. More specifically, there was no information regarding students' cognitive factors, students' learning environment, students' self-regulated learning characteristics and motivation, teaching and assessment practices, and characteristics of students' job. Therefore, essential patterns of information that could have helped to unravel the complex nature of academic performance in higher education could not be identified. Undoubtedly, the quality of the carried-out classification, regression, and clustering analyses would have improved if such information had been available.

An additional limitation is the empirical design of the ANNs. Many design choices need to be made when designing ANNs, namely, number of layers, number of neurons per layer, the activation function of each layer, among others. Even though this doctoral dissertation followed a

systematic process to identify the 'best' possible hyper-parameters of each ANN (e.g., learning rate, momentum, overlapping factor), some specific choices were made based on a criterion that would maximize interpretability. For example, the number of layers and its corresponding number of neurons in the MLP, and the size of the grid of the SOM. As these parameters also influence the performance of an ANN, it could be the case that a different architecture would have led to better results.

A final limitation is the way in which overfitting of predictive systems based on ANNs was addressed in this research. Overfitting occurs when an ANN shows good performance during the training phase but fails to identify unseen cases during the testing phase because the ANN has also identified noisy patterns in the data set, and it is trying to "fit" this noise rather than only the information contained in the patterns. In this doctoral dissertation, the so-called training and testing split of the data set was used in order to avoid overfitting. Therefore, it was possible to verify how similar the quality measures were between the testing and the training data sets. As such, any considerable decrease in the quality measures using the testing data set would have suggested overfitting. Nevertheless, there are more sophisticated ways to avoid overfitting, such as the use of a cross-validation set and regularization techniques. Although such a selection might have resulted in an improvement of the predictive models discussed in studies 2 and 3, it was also a technical complex issue beyond the aim of this doctoral dissertation.

Future research

Future research on the analysis of academic performance could focus on several topics. The first research topic is investigating how to create SES indices effectively. Although several statistical techniques to obtain such indices exist (e.g., factor analysis, principal component analysis), their amount of explained variance can still be increased. As such, future research could

explore statistical techniques such as categorical principal component analysis (CatPCA), which genuinely recognizes the categorical nature of several SES indicators.

The second research topic is to use ANNs in predictive validity studies in higher education. It has been rather common to determine the predictive validity of a presumed set of predictors through classical statistical techniques. Thus, in most cases, a model that only includes significant predictors is kept for further analysis. However, the significance level is a tricky concept highly dependent on additional factors not related to academic performance, such as the sample size. In contrast, predictive models based on ANNs could efficiently show the predictive validity of a set of predictors, whether they are significant or not.

The third research topic is to investigate the use of different ANNs types, related to statistical approaches not considered in this doctoral dissertation, such as the Bayesian neural networks (BNNs). The main reason behind this suggestion is that BNNs make it possible to combine ANNs and stochastic modeling. BNNs treat the training process as a probabilistic inference problem. A BNN analysis also allows the comparison of different models (e.g., MLP, RBF) using only the training data. Besides, BNNs provide a structured framework to tackle the issue of model complexity and avoid problems related to the use of maximum likelihood estimation (Bishop, 1997).

The fourth research topic is to deepen the use of unsupervised learning paradigms. By doing so, the educational sciences field would benefit from the development of *real* predictive systems that do not require information on the output variable for the prediction. This type of predictive systems could help higher education institutions not only to optimize the admission process but also to develop early prediction systems that identify students who are at risk of

academic failure. Then, the efforts to develop intervention strategies to support these students could be better allocated.

The fifth topic is to increase the use of deep learning techniques. Deep learning refers to neural networks with several layers of neurons. In their recent contribution, Hernández-Blanco, Herrera-Flores, Tomás, and Navarro-Colorado (2019) point out that the use of deep learning has benefit the task of predicting students' academic performance within EDM, by outperforming the traditional machine learning techniques. Also, Hernández-Blanco et al. (2019) have suggested that the advances in deep learning would increase the theoretical understanding and the interpretability of the resulting predictive models of academic performance.

This doctoral dissertation has contributed to the conceptual and methodological understanding of the ANNs when used to analyze academic performance in higher education. In this respect, three different topologies (MLP, RBFNN, and SOM) framed into two learning paradigms (supervised and unsupervised) were conceptualized and used to perform three specific tasks, namely, classification, prediction, and clustering. Consequently, the relationship between academic performance in higher education and several of its predictors could be analyzed in an accurate and interpretable way. Nevertheless, it is important to note that the findings of this doctoral dissertation are not suggesting dismissing the use of traditional statistical techniques when analyzing educational data. Multiple linear regression, logistic regression, factor analysis, among other well-known and widely used techniques, have contributed to identifying which are the determinants of academic performance. What it is recommended, instead, is to recognize ANNs as a powerful alternative method, inside the analysis toolbox of the quantitative educational researcher, so they can be used to precisely identify unknown patterns existing within vast volumes of educational data. Ultimately, students, teachers, and institutions could take advantage of a more

precise and more in-depth understanding of the factors related to academic performance in higher education.

References

Astin, A. W. (1993). *What matters in college? Four years revisited*. San Francisco, USA: Jossey-Bass.

Bishop, C. (1997). Bayesian principal component analysis. Retrieved from: https://papers.nips.cc/paper/1549-bayesian-pca.pdf

Hawking, S., & Mlodinow, L. (2010). *The grand design*. New York: Bantam Books.

Hernández-Blanco, A., Herrera-Flores, B., Tomás, D., & Navarro-Colorado, B. (2019). A systematic review of deep learning approaches to educational data mining. *Complexity, 1*, 1-22. doi:

Hodges, B., Bedford, T., Hartley, J., Klinger, C., Murray, N., O'Rourke, J., & Schofield, N. (2013). *Enabling retention: processes and strategies for improving student retention in university-based enabling programs: Final report 2013*. Sydney, Australia: Australian Government Office for Learning and Teaching.

Pitman, T., Trinidad, S., Devlin, M., Harvey, A., Brett, M., & McKay, J. (2016). *Pathways to higher education: The efficacy of enabling and sub-bachelor pathways for disadvantaged students*. National Centre for Student Equity in Higher Education (NCSEHE), Perth: Curtin University.

Sirin, S. (2005). Socioeconomic status and academic achievement: A meta-analytic review of research. *Review of Educational Research, 75*, 417-453. doi:10.3102/00346543075003417

Smith, E. (2016). Can higher education compensate for society? Modelling the determinants of academic success at university. *British Journal of Sociology of Education, 37,* 970-992. doi:10.1080/01425692.2014.987728

Van Ewijk, R., & Sleegers, P. (2010). The effect of peer socioeconomic status on student achievement: A meta-analysis. *Educational Research Review, 5,* 134-150. doi:10.1016/j.edurev.2010.02.001

Wang, H., Kong, M., Shan, W., & Vong, S. K. (2010). The effects of doing parttime jobs on college student academic performance and social life in a Chinese society. *Journal of Education and Work, 23,* 79-94. doi:10.1080/13639080903418402

Annex 1. Descriptive statistics of the analyzed predictors

Tuition fees (loan)

	Frequency	Percent	Valid Percent	Cumulative Percent
No	111190	68.6	68.6	68.6
Yes	50840	31.4	31.4	100.0
Total	162030	100.0	100.0	

Tuition fees (own resources)

	Frequency	Percent	Valid Percent	Cumulative Percent
No	114349	70.6	70.6	70.6
Yes	47681	29.4	29.4	100.0
Total	162030	100.0	100.0	

Tuition fees (parents)

	Frequency	Percent	Valid Percent	Cumulative Percent
No	48065	29.7	29.7	29.7
Yes	113965	70.3	70.3	100.0
Total	162030	100.0	100.0	

Tuition fees (scholarship)

	Frequency	Percent	Valid Percent	Cumulative Percent
No	139343	86.0	86.0	86.0
Yes	22687	14.0	14.0	100.0
Total	162030	100.0	100.0	

Father's educational level

		Frequency	Percent	Valid Percent	Cumulative Percent
	Bachelor's degree completed	27654	17.1	17.1	17.1
	Bachelor's degree incomplete	5653	3.5	3.5	20.6
	Technical or technological degree completed	16392	10.1	10.1	30.7
	Technical or technological degree incomplete	3957	2.4	2.4	33.1
	No education	3641	2.2	2.2	35.4
	Unknown	3568	2.2	2.2	37.6
	Postgraduate education	14019	8.7	8.7	46.2
	Primary education completed	13590	8.4	8.4	54.6
	Primary education incomplete	19544	12.1	12.1	66.7
	Secondary education completed	36877	22.8	22.8	89.4
	Secondary education incomplete	17135	10.6	10.6	100.0
	Total	162030	100.0	100.0	

Father's occupation

		Frequency	Percent	Valid Percent	Cumulative Percent
	Unemployed	4657	2.9	2.9	2.9
	Domestic worker	645	.4	.4	3.3
	Day laborer	2566	1.6	1.6	4.9
	N.A.	14017	8.7	8.7	13.5
	Employee at private sector	36527	22.5	22.5	36.1
	Public servant	12197	7.5	7.5	43.6
	Other activity or occupation	17065	10.5	10.5	54.1
	Employer	13146	8.1	8.1	62.2
	Retired	14705	9.1	9.1	71.3
	Unsalaried worker at family business	367	.2	.2	71.5
	Self-employed	45370	28.0	28.0	99.5
	Unsalaried worker at non-family business	768	.5	.5	100.0
	Total	162030	100.0	100.0	

Mother's educational level

		Frequency	Percent	Valid Percent	Cumulative Percent
	Bachelor's degree completed	28207	17.4	17.4	17.4
	Bachelor's degree incomplete	5364	3.3	3.3	20.7
	Technical or technological degree completed	23116	14.3	14.3	35.0
	Technical or technological degree incomplete	4848	3.0	3.0	38.0
	None	1281	.8	.8	38.8
	Unknown	615	.4	.4	39.1
	Postgraduate education	13574	8.4	8.4	47.5
	Primary education completed	12107	7.5	7.5	55.0
	Primary education incomplete	14053	8.7	8.7	63.7
	Secondary education completed	40066	24.7	24.7	88.4
	Secondary education incomplete	18799	11.6	11.6	100.0
	Total	162030	100.0	100.0	

Mother's occupation

		Frequency	Percent	Valid Percent	Cumulative Percent
	Unemployed	15251	9.4	9.4	9.4
	Domestic worker	14234	8.8	8.8	18.2
	Day laborer	290	.2	.2	18.4
	N.A.	12805	7.9	7.9	26.3
	Employee at private sector	24447	15.1	15.1	41.4
	Public servant	17967	11.1	11.1	52.5
	Other activity or occupation	16302	10.1	10.1	62.5
	Employer	6656	4.1	4.1	66.6
	Retired	8532	5.3	5.3	71.9
	Unsalaried worker at family business	10205	6.3	6.3	78.2
	Self-employed	34376	21.2	21.2	99.4
	Unsalaried worker at non-family business	965	.6	.6	100.0
	Total	162030	100.0	100.0	

Internet access

		Frequency	Percent	Valid Percent	Cumulative Percent
	No	22742	14.0	14.0	14.0
	Yes	139288	86.0	86.0	100.0
	Total	162030	100.0	100.0	

PC at home

		Frequency	Percent	Valid Percent	Cumulative Percent
	No	11267	7.0	7.0	7.0
	Yes	150763	93.0	93.0	100.0
	Total	162030	100.0	100.0	

Type of accommodation

		Frequency	Percent	Valid Percent	Cumulative Percent
	Permanent accommodation	128017	79.0	79.0	79.0
	Temporal accommodation	34013	21.0	21.0	100.0
	Total	162030	100.0	100.0	

Students' house stratum

		Frequency	Percent	Valid Percent	Cumulative Percent
	0	445	.3	.3	.3
	1	21787	13.4	13.4	13.7
	2	51518	31.8	31.8	45.5
	3	54490	33.6	33.6	79.1
	4	21559	13.3	13.3	92.5
	5	8051	5.0	5.0	97.4
	6	4180	2.6	2.6	100.0
	Total	162030	100.0	100.0	

Head of household

		Frequency	Percent	Valid Percent	Cumulative Percent
	No	151247	93.3	93.3	93.3
	Yes	10783	6.7	6.7	100.0
	Total	162030	100.0	100.0	

Students' gender

		Frequency	Percent	Valid Percent	Cumulative Percent
	Female	97030	59.9	59.9	59.9
	Male	65000	40.1	40.1	100.0
	Total	162030	100.0	100.0	

High school type

		Frequency	Percent	Valid Percent	Cumulative Percent
	Private	70322	43.4	43.4	43.4
	Public	91708	56.6	56.6	100.0
	Total	162030	100.0	100.0	

High school academic calendar

		Frequency	Percent	Valid Percent	Cumulative Percent
	A (starts in january)	144289	89.1	89.1	89.1
	B (starts in september)	10037	6.2	6.2	95.2
	Flexible	7704	4.8	4.8	100.0
	Total	162030	100.0	100.0	

High school diploma

		Frequency	Percent	Valid Percent	Cumulative Percent
	Academic	99797	61.6	61.6	61.6
	Academic and Technical	29276	18.1	18.1	79.7
	Unknown	50	.0	.0	79.7
	Teacher-training	5869	3.6	3.6	83.3
	Technical	27038	16.7	16.7	100.0
	Total	162030	100.0	100.0	

High school "gender"

		Frequency	Percent	Valid Percent	Cumulative Percent
	Girls-only	19840	12.2	12.2	12.2
	Boys-only	5089	3.1	3.1	15.4
	Mixed	137101	84.6	84.6	100.0
	Total	162030	100.0	100.0	

High school schedule

		Frequency	Percent	Valid Percent	Cumulative Percent
	Ordinary	55066	34.0	34.0	34.0
	Morning-only	81670	50.4	50.4	84.4
	Evening-only	2983	1.8	1.8	86.2
	Weekend-only	1591	1.0	1.0	87.2
	Afternoon-only	20695	12.8	12.8	100.0
	Full day	25	.02	.02	100.0
	Total	162030	100.0	100.0	

Working hours per week

		Frequency	Percent	Valid Percent	Cumulative Percent
	0	64044	39.5	39.5	39.5
	Between 11 and 20	18222	11.2	11.2	50.8
	Between 21 and 30	13294	8.2	8.2	59.0
	More than 30	45506	28.1	28.1	87.1
	Less than 10	20964	12.9	12.9	100.0
	Total	162030	100.0	100.0	

Academic program

		Frequency	Percent	Valid Percent	Cumulative Percent
	Agronomy, veterinary and related degrees	2148	1.3	1.3	1.3
	Arts	5724	3.5	3.5	4.9
	Education	14161	8.7	8.7	13.6
	Health	18534	11.4	11.4	25.0
	Social and human sciences	34036	21.0	21.0	46.0
	Economics, management and accounting	41096	25.4	25.4	71.4
	Engineering, architecture, urban planning and related degrees	41469	25.6	25.6	97.0
	Mathematics and natural sciences	3435	2.1	2.1	99.1
	Not specified	1427	.9	.9	100.0
	Total	162030	100.0	100.0	

University type

		Frequency	Percent	Valid Percent	Cumulative Percent
	Private (corporation funding)	53212	32.8	32.8	32.8
	Private (foundation funding)	51118	31.5	31.5	64.4
	Public (departmental funding)	27669	17.1	17.1	81.5
	Public (municipalities funding)	3130	1.9	1.9	83.4
	Public (national funding)	26877	16.6	16.6	100.0
	Public (special regime)	24	.015	.015	100.0
	Total	162030	100.0	100.0	

	N	Mean	Std. Deviation
Quantitative Reasoning score	162030	155.80	30.17
Age at SABERPRO	162030	23.50	2.89
Monthly family income	162030	2.88	1.49
Number of persons in charge	162030	0.29	0.81
Number of rooms in the house	162030	2.85	0.99
Number of semesters at SABER PRO	162030	9.30	1.10
SABER 11-mathematics score	162030	52.97	11.42
SABER 11-Spanish score	162030	52.57	8.86
SABER11-biology score	162030	51.04	8.93
SABER11-chemistry score	162030	50.53	8.80
SABER11-philosophy score	162030	47.80	10.11
SABER11-physics score	162030	49.04	9.70
SABER11-social sciences score	162030	51.49	9.22

Annex 2. Training of the MLPs to classify high-performance group

Parameters of the model			Input rates				Name of the ANN quality indicator				
							Sensitivity/Recall	Specificity	Precision	Accuracy	F1 Score
							TP/(TP+FN)	TN/(TN+FP)	TP/(TP+FP)	(TP+TN)/(TP+FP+FN+TN)	2TP/(2TP+FP+FN)
Output function	Learning rate	Momentum	TP	TN	FP	FN	TPR	TNR / SPC	PPV	ACC	F1
Sigmoid	0.1	0.1	1985	22315	7831	275	0.878	0.740	0.202	0.750	0.329
		0.2	2017	21911	8235	243	0.892	0.727	0.197	0.738	0.322
		0.3	2071	21060	9086	189	0.916	0.699	0.186	0.714	0.309
		0.4	2086	20792	9354	174	0.923	0.690	0.182	0.706	0.305
		0.5	2161	18063	12083	99	0.956	0.599	0.152	0.624	0.262
		0.6	2059	20591	9555	201	0.911	0.683	0.177	0.699	0.297
		0.7	1811	24646	5500	449	0.801	0.818	0.248	0.816	0.378
		0.8	1914	22498	7648	346	0.847	0.746	0.200	0.753	0.324
		0.9	2057	19939	10207	203	0.910	0.661	0.168	0.679	0.283
	0.2	0.1	2048	20892	9254	212	0.906	0.693	0.181	0.708	0.302
		0.2	2105	19853	10293	155	0.931	0.659	0.170	0.678	0.287
		0.3	2025	21614	8532	235	0.896	0.717	0.192	0.729	0.316
		0.4	2149	18387	11759	111	0.951	0.610	0.155	0.634	0.266
		0.5	2138	18536	11610	122	0.946	0.615	0.156	0.638	0.267
		0.6	2172	17327	12819	88	0.961	0.575	0.145	0.602	0.252
		0.7	2115	18427	11719	145	0.936	0.611	0.153	0.634	0.263
		0.8	1919	20740	9406	341	0.849	0.688	0.169	0.699	0.283
		0.9	1901	22651	7495	359	0.841	0.751	0.202	0.758	0.326
	0.3	0.1	2105	20025	10121	155	0.931	0.664	0.172	0.683	0.291
		0.2	2127	19441	10705	133	0.941	0.645	0.166	0.666	0.282
		0.3	2153	18353	11793	107	0.953	0.609	0.154	0.633	0.266
		0.4	2055	20562	9584	205	0.909	0.682	0.177	0.698	0.296
		0.5	2137	17520	12626	123	0.946	0.581	0.145	0.607	0.251
		0.6	2051	20401	9745	209	0.908	0.677	0.174	0.693	0.292
		0.7	2044	20101	10045	216	0.904	0.667	0.169	0.683	0.285
		0.8	2101	19067	11079	159	0.930	0.632	0.159	0.653	0.272
		0.9	1834	23434	6712	426	0.812	0.777	0.215	0.780	0.339
	0.4	0.1	2067	20898	9248	193	0.915	0.693	0.183	0.709	0.305
		0.2	2143	17940	12206	117	0.948	0.595	0.149	0.620	0.258
		0.3	1973	22163	7983	287	0.873	0.735	0.198	0.745	0.323
		0.4	2091	18990	11156	169	0.925	0.630	0.158	0.651	0.270
		0.5	1983	21111	9035	277	0.877	0.700	0.180	0.713	0.299
		0.6	2141	16966	13180	119	0.947	0.563	0.140	0.590	0.244
		0.7	1925	21218	8928	335	0.852	0.704	0.177	0.714	0.294
		0.8	1900	22832	7314	360	0.841	0.757	0.206	0.763	0.331
		0.9	1800	23948	6198	460	0.796	0.794	0.225	0.795	0.351
	0.5	0.1	2087	19614	10532	173	0.923	0.651	0.165	0.670	0.281

		0.2	2086	20272	9874	174	0.923	0.672	0.174	0.690	0.293
		0.3	2037	20656	9490	223	0.901	0.685	0.177	0.700	0.295
		0.4	2076	19308	10838	184	0.919	0.640	0.161	0.660	0.274
		0.5	2073	18648	11498	187	0.917	0.619	0.153	0.639	0.262
		0.6	1774	22651	7495	486	0.785	0.751	0.191	0.754	0.308
		0.7	2113	18389	11757	147	0.935	0.610	0.152	0.633	0.262
		0.8	2065	18548	11598	195	0.914	0.615	0.151	0.636	0.259
		0.9	2061	19311	10835	199	0.912	0.641	0.160	0.660	0.272
		0.1	1976	21825	8321	284	0.874	0.724	0.192	0.734	0.315
		0.2	2060	19833	10313	200	0.912	0.658	0.166	0.676	0.282
		0.3	1730	24047	6099	530	0.765	0.798	0.221	0.795	0.343
		0.4	1828	24253	5893	432	0.809	0.805	0.237	0.805	0.366
	0.1	0.5	1922	23246	6900	338	0.850	0.771	0.218	0.777	0.347
		0.6	2118	19421	10725	142	0.937	0.644	0.165	0.665	0.280
		0.7	2056	21307	8839	204	0.910	0.707	0.189	0.721	0.313
		0.8	2170	16523	13623	90	0.960	0.548	0.137	0.577	0.240
		0.9	2085	19415	10731	175	0.923	0.644	0.163	0.663	0.277
		0.1	2048	20558	9588	212	0.906	0.682	0.176	0.698	0.295
		0.2	2142	19033	11113	118	0.948	0.631	0.162	0.653	0.276
		0.3	1991	22184	7692	269	0.881	0.743	0.206	0.752	0.333
		0.4	2040	21597	8549	220	0.903	0.716	0.193	0.729	0.318
	0.2	0.5	2094	19977	10169	166	0.927	0.663	0.171	0.681	0.288
		0.6	1807	24683	5463	453	0.800	0.819	0.249	0.817	0.379
		0.7	1807	24331	5815	453	0.800	0.807	0.237	0.807	0.366
		0.8	2231	10091	20055	29	0.987	0.335	0.100	0.380	0.182
Linear sigmoid		0.9	2246	6186	23960	14	0.994	0.205	0.086	0.260	0.158
		0.1	2136	18887	11259	124	0.945	0.627	0.159	0.649	0.273
		0.2	2162	18260	11886	98	0.957	0.606	0.154	0.630	0.265
		0.3	2035	21515	8631	225	0.900	0.714	0.191	0.727	0.315
		0.4	2102	19441	10705	158	0.930	0.645	0.164	0.665	0.279
	0.3	0.5	2131	17916	12230	129	0.943	0.594	0.148	0.619	0.256
		0.6	2194	14515	15631	66	0.971	0.481	0.123	0.516	0.218
		0.7	2206	11747	18399	54	0.976	0.390	0.107	0.431	0.193
		0.8	2081	17861	12285	179	0.921	0.592	0.145	0.615	0.250
		0.9	2168	14742	15404	92	0.959	0.489	0.123	0.522	0.219
		0.1	2088	18913	11233	172	0.924	0.627	0.157	0.648	0.268
		0.2	1792	25048	5098	468	0.793	0.831	0.260	0.828	0.392
		0.3	2245	10268	19878	15	0.993	0.341	0.101	0.386	0.184
		0.4	1966	22359	7787	294	0.870	0.742	0.202	0.751	0.327
	0.4	0.5	2229	12215	17931	31	0.986	0.405	0.111	0.446	0.199
		0.6	2234	8477	21669	26	0.988	0.281	0.093	0.331	0.171
		0.7	2068	18617	11529	192	0.915	0.618	0.152	0.638	0.261
		0.8	1835	21805	8341	425	0.812	0.723	0.180	0.729	0.295
		0.9	2092	17466	12680	168	0.926	0.579	0.142	0.604	0.246

		0.1	2128	18793	11353	132	0.942	0.623	0.158	0.646	0.270
		0.2	2026	21137	9009	234	0.896	0.701	0.184	0.715	0.305
		0.3	2061	20800	9346	199	0.912	0.690	0.181	0.705	0.302
		0.4	2074	19557	10589	186	0.918	0.649	0.164	0.667	0.278
	0.5	0.5	2098	17617	12529	162	0.928	0.584	0.143	0.608	0.248
		0.6	1854	23054	7092	406	0.820	0.765	0.207	0.769	0.331
		0.7	2034	20024	10122	226	0.900	0.664	0.167	0.681	0.282
		0.8	2113	15605	14541	147	0.935	0.518	0.127	0.547	0.223
		0.9	2257	367	29779	3	0.999	0.012	0.070	0.081	0.132
		0.1	237	28575	1571	2023	0.105	0.948	0.131	0.889	0.117
		0.2	1974	18886	11260	286	0.873	0.626	0.149	0.644	0.255
		0.3	1940	10519	19627	320	0.858	0.349	0.090	0.384	0.163
		0.4	1727	22777	7369	533	0.764	0.756	0.190	0.756	0.304
	0.1	0.5	1707	22962	7184	553	0.755	0.762	0.192	0.761	0.306
		0.6	1937	19276	10870	323	0.857	0.639	0.151	0.655	0.257
		0.7	1525	25355	4791	735	0.675	0.841	0.241	0.829	0.356
		0.8	1743	21883	8263	517	0.771	0.726	0.174	0.729	0.284
		0.9	1788	24090	6056	472	0.791	0.799	0.228	0.799	0.354
		0.1	1329	24051	6095	931	0.588	0.798	0.179	0.783	0.274
		0.2	1797	23026	7120	463	0.795	0.764	0.202	0.766	0.322
		0.3	1771	22681	7465	489	0.784	0.752	0.192	0.755	0.308
		0.4	1869	22249	7897	391	0.827	0.738	0.191	0.744	0.311
	0.2	0.5	1565	25137	5009	695	0.692	0.834	0.238	0.824	0.354
		0.6	1742	22812	7334	518	0.771	0.757	0.192	0.758	0.307
		0.7	2030	18388	11758	230	0.898	0.610	0.147	0.630	0.253
		0.8	1820	20439	9707	440	0.805	0.678	0.158	0.687	0.264
Softmax		0.9	1892	22298	7848	368	0.837	0.740	0.194	0.746	0.315
		0.1	1920	20696	9450	340	0.850	0.687	0.169	0.698	0.282
		0.2	1647	24377	5769	613	0.729	0.809	0.222	0.803	0.340
		0.3	1623	24641	5505	637	0.718	0.817	0.228	0.810	0.346
		0.4	1903	20977	9169	357	0.842	0.696	0.172	0.706	0.285
	0.3	0.5	1667	24348	5798	593	0.738	0.808	0.223	0.803	0.343
		0.6	1948	19702	10444	312	0.862	0.654	0.157	0.668	0.266
		0.7	1388	25134	5012	872	0.614	0.834	0.217	0.818	0.321
		0.8	2003	19276	10870	257	0.886	0.639	0.156	0.657	0.265
		0.9	1943	15517	14629	317	0.860	0.515	0.117	0.539	0.206
		0.1	1959	17878	12268	301	0.867	0.593	0.138	0.612	0.238
		0.2	1445	24207	5939	815	0.639	0.803	0.196	0.792	0.300
		0.3	1860	22912	7234	400	0.823	0.760	0.205	0.764	0.328
	0.4	0.4	1649	23175	6971	611	0.730	0.769	0.191	0.766	0.303
		0.5	1906	19951	10195	354	0.843	0.662	0.158	0.674	0.265
		0.6	1987	17649	12497	273	0.879	0.585	0.137	0.606	0.237
		0.7	1859	20684	9462	401	0.823	0.686	0.164	0.696	0.274
		0.8	1905	22533	7613	355	0.843	0.747	0.200	0.754	0.323

	0.9	1932	20768	9378	328	0.855	0.689	0.171	0.700	0.285
	0.1	1531	25793	4353	729	0.677	0.856	0.260	0.843	0.376
	0.2	1888	19599	10547	372	0.835	0.650	0.152	0.663	0.257
	0.3	1936	17805	12341	324	0.857	0.591	0.136	0.609	0.234
	0.4	1854	19478	10668	406	0.820	0.646	0.148	0.658	0.251
0.5	0.5	1671	21840	8306	589	0.739	0.724	0.167	0.726	0.273
	0.6	2009	19492	10654	251	0.889	0.647	0.159	0.663	0.269
	0.7	1925	20369	9777	335	0.852	0.676	0.165	0.688	0.276
	0.8	1759	23992	6154	501	0.778	0.796	0.222	0.795	0.346
	0.9	1826	24359	5787	434	0.808	0.808	0.240	0.808	0.370

Annex 3. Training of the MLPs to classify low-performance group

Parameters of the model			Input rates				Name of the ANN quality indicator				
							Sensitivity/Recall	Specificity	Precision	Accuracy	F1 Score
							TP/(TP+FN)	TN/(TN+FP)	TP/(TP+FP)	(TP+TN)/(TP+FP+FN+TN)	2TP/(2TP+FP+FN)
Output function	**Learning rate**	**Momentum**	**TP**	**TN**	**FP**	**FN**	**TPR**	**TNR / SPC**	**PPV**	**ACC**	**F1**
		0.1	5000	18076	7750	1580	0.760	0.700	0.392	0.712	0.517
		0.2	5091	17701	8125	1489	0.774	0.685	0.385	0.703	0.514
		0.3	5026	17678	8148	1554	0.764	0.685	0.382	0.701	0.509
		0.4	4901	18875	6951	1679	0.745	0.731	0.414	0.734	0.532
	0.1	0.5	5163	17735	8091	1417	0.785	0.687	0.390	0.707	0.521
		0.6	5417	16654	9172	1163	0.823	0.645	0.371	0.681	0.512
		0.7	5345	17037	8789	1235	0.812	0.660	0.378	0.691	0.516
		0.8	5462	16339	9487	1118	0.830	0.633	0.365	0.673	0.507
		0.9	5660	14141	11685	920	0.860	0.548	0.326	0.611	0.473
		0.1	4916	18217	7609	1664	0.747	0.705	0.392	0.714	0.515
		0.2	4823	19484	6342	1757	0.733	0.754	0.432	0.750	0.544
		0.3	4903	18914	6912	1677	0.745	0.732	0.415	0.735	0.533
		0.4	5013	17716	8110	1567	0.762	0.686	0.382	0.701	0.509
	0.2	0.5	4731	19044	6782	1849	0.719	0.737	0.411	0.734	0.523
		0.6	4057	20814	5012	2523	0.617	0.806	0.447	0.767	0.518
		0.7	4425	19673	6153	2155	0.672	0.762	0.418	0.744	0.516
		0.8	5582	14001	11825	998	0.848	0.542	0.321	0.604	0.465
		0.9	5009	17668	8158	1571	0.761	0.684	0.380	0.700	0.507
Sigmoid		0.1	5403	16583	9243	1177	0.821	0.642	0.369	0.678	0.509
		0.2	4442	20449	5377	2138	0.675	0.792	0.452	0.768	0.542
		0.3	4899	18463	7363	1681	0.745	0.715	0.400	0.721	0.520
		0.4	5445	15522	10304	1135	0.828	0.601	0.346	0.647	0.488
	0.3	0.5	5032	18000	7826	1548	0.765	0.697	0.391	0.711	0.518
		0.6	5307	16599	9227	1273	0.807	0.643	0.365	0.676	0.503
		0.7	5258	16095	9731	1322	0.799	0.623	0.351	0.659	0.488
		0.8	6015	10151	15675	565	0.914	0.393	0.277	0.499	0.426
		0.9	5479	16028	9798	1101	0.833	0.621	0.359	0.664	0.501
		0.1	4167	21009	4817	2413	0.633	0.813	0.464	0.777	0.535
		0.2	3972	21697	4129	2608	0.604	0.840	0.490	0.792	0.541
		0.3	5230	16459	9367	1350	0.795	0.637	0.358	0.669	0.494
		0.4	4832	17869	7957	1748	0.734	0.692	0.378	0.701	0.499
	0.4	0.5	4036	20845	4981	2544	0.613	0.807	0.448	0.768	0.518
		0.6	3953	20964	4862	2627	0.601	0.812	0.448	0.769	0.514
		0.7	4565	17442	8384	2015	0.694	0.675	0.353	0.679	0.468
		0.8	3975	20293	5533	2605	0.604	0.786	0.418	0.749	0.494
		0.9	4888	17234	8592	1692	0.743	0.667	0.363	0.683	0.487
	0.5	0.1	4590	19512	6314	1990	0.698	0.756	0.421	0.744	0.525

		0.2	4529	19538	6288	2051	0.688	0.757	0.419	0.743	0.521
		0.3	3769	21779	4047	2811	0.573	0.843	0.482	0.788	0.524
		0.4	3840	21324	4502	2740	0.584	0.826	0.460	0.777	0.515
		0.5	4523	17920	7906	2057	0.687	0.694	0.364	0.693	0.476
		0.6	4872	17178	8648	1708	0.740	0.665	0.360	0.680	0.485
		0.7	4911	17053	8773	1669	0.746	0.660	0.359	0.678	0.485
		0.8	3317	22143	3683	3263	0.504	0.857	0.474	0.786	0.489
		0.9	4305	20101	5725	2275	0.654	0.778	0.429	0.753	0.518
		0.1	4797	19118	6708	1783	0.729	0.740	0.417	0.738	0.530
		0.2	5037	17937	7889	1543	0.766	0.695	0.390	0.709	0.516
		0.3	4021	20532	5294	2559	0.611	0.795	0.432	0.758	0.506
		0.4	5089	17114	8712	1491	0.773	0.663	0.369	0.685	0.499
	0.1	0.5	5436	16447	9379	1144	0.826	0.637	0.367	0.675	0.508
		0.6	3685	21501	4325	2895	0.560	0.833	0.460	0.777	0.505
		0.7	4749	18534	7292	1831	0.722	0.718	0.394	0.718	0.510
		0.8	5477	16639	9187	1103	0.832	0.644	0.373	0.682	0.516
		0.9	6107	11360	14466	473	0.928	0.440	0.297	0.539	0.450
		0.1	5421	15842	9984	1159	0.824	0.613	0.352	0.656	0.493
		0.2	5865	13140	12686	715	0.891	0.509	0.316	0.586	0.467
		0.3	5347	16047	9779	1233	0.813	0.621	0.353	0.660	0.493
		0.4	4914	18823	7003	1666	0.747	0.729	0.412	0.732	0.531
	0.2	0.5	6003	12422	13404	577	0.912	0.481	0.309	0.569	0.462
		0.6	6306	9470	16356	274	0.958	0.367	0.278	0.487	0.431
		0.7	4801	19409	6417	1779	0.730	0.752	0.428	0.747	0.539
		0.8	6109	10167	15659	471	0.928	0.394	0.281	0.502	0.431
Linear sigmoid		0.9	4247	20138	5688	2333	0.645	0.780	0.427	0.752	0.514
		0.1	3480	22341	3485	3100	0.529	0.865	0.500	0.797	0.514
		0.2	5815	14450	11376	765	0.884	0.560	0.338	0.625	0.489
		0.3	4365	19584	6242	2215	0.663	0.758	0.412	0.739	0.508
		0.4	5866	14131	11695	714	0.891	0.547	0.334	0.617	0.486
	0.3	0.5	4217	20956	4870	2363	0.641	0.811	0.464	0.777	0.538
		0.6	4455	19714	6112	2125	0.677	0.763	0.422	0.746	0.520
		0.7	6373	7422	18404	207	0.969	0.287	0.257	0.426	0.406
		0.8	4342	19764	6062	2238	0.660	0.765	0.417	0.744	0.511
		0.9	6261	7991	17835	319	0.952	0.309	0.260	0.440	0.408
		0.1	6187	10968	14858	393	0.940	0.425	0.294	0.529	0.448
		0.2	4947	18410	7416	1633	0.752	0.713	0.400	0.721	0.522
		0.3	3170	21264	4562	3410	0.482	0.823	0.410	0.754	0.443
		0.4	6145	10793	15033	435	0.934	0.418	0.290	0.523	0.443
	0.4	0.5	3583	21760	4066	2997	0.545	0.843	0.468	0.782	0.504
		0.6	5355	17086	8740	1225	0.814	0.662	0.380	0.692	0.518
		0.7	4402	19648	6178	2178	0.669	0.761	0.416	0.742	0.513
		0.8	6072	10444	15382	508	0.923	0.404	0.283	0.510	0.433
		0.9	6393	6556	19270	187	0.972	0.254	0.249	0.400	0.397

		0.1	4551	19539	6287	2029	0.692	0.757	0.420	0.743	0.523
		0.2	5746	14607	11219	834	0.873	0.566	0.339	0.628	0.488
		0.3	6467	4911	20915	113	0.983	0.190	0.236	0.351	0.381
		0.4	4581	19985	5841	1999	0.696	0.774	0.440	0.758	0.539
	0.5	0.5	5903	12667	13159	677	0.897	0.490	0.310	0.573	0.460
		0.6	5865	12133	13693	715	0.891	0.470	0.300	0.555	0.449
		0.7	6135	10496	15330	445	0.932	0.406	0.286	0.513	0.438
		0.8	5846	11821	14005	734	0.888	0.458	0.294	0.545	0.442
		0.9	6472	4814	21012	108	0.984	0.186	0.235	0.348	0.380
		0.1	5234	6480	19346	1346	0.795	0.251	0.213	0.361	0.336
		0.2	6289	819	25007	291	0.956	0.032	0.201	0.219	0.332
		0.3	2019	16596	9230	4561	0.307	0.643	0.179	0.574	0.226
		0.4	571	24322	1504	6009	0.087	0.942	0.275	0.768	0.132
	0.1	0.5	2468	20157	5669	4112	0.375	0.780	0.303	0.698	0.335
		0.6	3671	18887	6939	2909	0.558	0.731	0.346	0.696	0.427
		0.7	4790	10949	14877	1790	0.728	0.424	0.244	0.486	0.365
		0.8	5302	11677	14149	1278	0.806	0.452	0.273	0.524	0.407
		0.9	6393	1608	24218	187	0.972	0.062	0.209	0.247	0.344
		0.1	536	25042	784	6044	0.081	0.970	0.406	0.789	0.136
		0.2	2520	21468	4358	4060	0.383	0.831	0.366	0.740	0.374
		0.3	2925	21476	4350	3655	0.445	0.832	0.402	0.753	0.422
		0.4	5079	15543	10283	1501	0.772	0.602	0.331	0.636	0.463
	0.2	0.5	5118	10279	15547	1462	0.778	0.398	0.248	0.475	0.376
		0.6	4779	16202	9624	1801	0.726	0.627	0.332	0.647	0.456
		0.7	5750	12341	13485	830	0.874	0.478	0.299	0.558	0.445
		0.8	4989	10729	15097	1591	0.758	0.415	0.248	0.485	0.374
Softmax		0.9	5283	6422	19404	1297	0.803	0.249	0.214	0.361	0.338
		0.1	4152	18273	7553	2428	0.631	0.708	0.355	0.692	0.454
		0.2	3824	19089	6737	2756	0.581	0.739	0.362	0.707	0.446
		0.3	2877	16303	9523	3703	0.437	0.631	0.232	0.592	0.303
		0.4	2712	19873	5953	3868	0.412	0.769	0.313	0.697	0.356
	0.3	0.5	4727	14427	11399	1853	0.718	0.559	0.293	0.591	0.416
		0.6	3717	13052	12774	2863	0.565	0.505	0.225	0.517	0.322
		0.7	5708	8182	17644	872	0.867	0.317	0.244	0.429	0.381
		0.8	3949	13595	12231	2631	0.600	0.526	0.244	0.541	0.347
		0.9	6006	7278	18548	574	0.913	0.282	0.245	0.410	0.386
		0.1	4277	15708	10118	2303	0.650	0.608	0.297	0.617	0.408
		0.2	5611	9819	16007	969	0.853	0.380	0.260	0.476	0.398
		0.3	5031	9714	16112	1549	0.765	0.376	0.238	0.455	0.363
		0.4	3433	19540	6286	3147	0.522	0.757	0.353	0.709	0.421
	0.4	0.5	4810	11349	14477	1770	0.731	0.439	0.249	0.499	0.372
		0.6	5700	7416	18410	880	0.866	0.287	0.236	0.405	0.371
		0.7	3979	10053	15773	2601	0.605	0.389	0.201	0.433	0.302
		0.8	3578	15378	10448	3002	0.544	0.595	0.255	0.585	0.347

	0.9	5570	9077	16749	1010	0.847	0.351	0.250	0.452	0.385
	0.1	3726	18458	7368	2854	0.566	0.715	0.336	0.685	0.422
	0.2	3412	19528	6298	3168	0.519	0.756	0.351	0.708	0.419
	0.3	3712	18685	7141	2868	0.564	0.723	0.342	0.691	0.426
	0.4	4249	14584	11242	2331	0.646	0.565	0.274	0.581	0.385
0.5	0.5	5898	7346	18480	682	0.896	0.284	0.242	0.409	0.381
	0.6	4891	13548	12278	1689	0.743	0.525	0.285	0.569	0.412
	0.7	5532	8756	17070	1048	0.841	0.339	0.245	0.441	0.379
	0.8	4854	7062	18764	1726	0.738	0.273	0.206	0.368	0.321
	0.9	5117	8765	17041	1463	0.778	0.340	0.231	0.429	0.356

Annex 4[4]

Types of Gaussian functions

Ordinary RBF (ORBF): This type uses the exp activation function, so the activation of the RBF unit is a Gaussian "bump" as a function of the inputs. In ORBF, the Gaussian basis function takes form

$$\varphi_j\left(X\right)=\exp\left(-\sum_{p=1}^{P}\frac{1}{2\sigma_{jP}^2}\left(x_p-\mu_{jp}\right)^2\right)$$

Determining Basis Functions (radial basis function algorithms)

The two-step clustering algorithm is used to find the RBF centers and widths. For each cluster, the mean and standard deviation for each scale variable and proportion of each category for each categorical variable are derived. Using the results from clustering, the center of the *jth* RBF is set as:

$$\mu_{jp}=\begin{cases}\bar{x}_{jp} & \text{if pth variable is scale}\\ \pi_{jp} & \text{if pth variable is a dummy variable of a categorical variable}\end{cases}$$

where X_{jp} is the *j*th cluster mean of the *p*th input variable if it is scale, and πjp is the proportion of the category of a categorical variable that the *p*th input variable corresponds to. The width of the *j*th RBF is set as

$$\sigma_{jp}=h^{1/2}\begin{cases}s_{jp} & \text{if pth variable is scale}\\ \sqrt{p_{jp}\left(1-p_{jp}\right)} & \text{if pth variable is a dummy variable of a categorical variable}\end{cases}$$

where *Sjp* is the *j*th cluster standard deviation of the *p*th variable and $h>0$ is the RBF overlapping factor that controls the amount of overlap among the RBFs.

[4] The text included in this annex was retrieved from http://desktop-5vfq00m:51220/help/index.jsp?topic=/com.ibm.spss.statistics.help/idh_idd_rbf_variables.htm

Sensitivity Analysis

For each predictor *p* and each input pattern *m*, compute:

$$d_{pm} = \max_{x_{p_1}, x_{p_2} \in S_p} \left|\left| \hat{Y}_{p_1}^{(m)} - \hat{Y}_{p_2}^{(m)} \right|\right|$$

where $Y^{(m)}pk$ is the predicted output vector (standardized if standardization of output variable is used in training) using ($x^{(m)}1$,..., $x^{(m)}p-1$, xpk, $x^{(m)}p+1$,..., $x^{(m)}P$) as its input, and Sp= {$x^{\min}p$, $x^{(2)}p$, $x^{(3)}p$, $x^{(4)}p$, $x^{\max}p$} for scale predictors and (1,0,...,0),(0,1,0,...,0),...,(0,0,...,1)} for categorical predictors.

Then compute:

$$d_p = \frac{1}{M} \sum_{m=1}^{M} d_{pm}$$

and normalize the d_p to sum to 1 and report these normalized values as the sensitivity values for the predictors. This is the average maximum amount we can expect the output to change based on changes in the *p*th predictor. The greater the sensitivity, the more we expect the output to change when the predictor changes.