

Predicting general academic performance and identifying the differential contribution of participating variables using artificial neural networks

Mariel F. Musso^{ab}, Eva Kyndt^{ac}, Eduardo C. Cascallar^{ad}, Filip Dochy^a

^aKatholieke Universiteit Leuven, Belgium

^bUniversidad Argentina de La Empresa, Argentina

^cUniversity of Antwerp, Belgium

^dAssessment Group International, USA / Belgium

Article received 8 March 2013 / revised 2 July 2013 / accepted 16 July 2013 / available online 27 August 2013

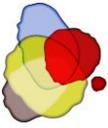
Abstract

Many studies have explored the contribution of different factors from diverse theoretical perspectives to the explanation of academic performance. These factors have been identified as having important implications not only for the study of learning processes, but also as tools for improving curriculum designs, tutorial systems, and students' outcomes. Some authors have suggested that traditional statistical methods do not always yield accurate predictions and/or classifications (Everson, 1995; Garson, 1998). This paper explores a relatively new methodological approach for the field of learning and education, but which is widely used in other areas, such as computational sciences, engineering and economics. This study uses cognitive and non-cognitive measures of students, together with background information, in order to design predictive models of student performance using artificial neural networks (ANN). These predictions of performance constitute a true predictive classification of academic performance over time, a year in advance of the actual observed measure of academic performance. A total sample of 864 university students of both genders, ages ranging between 18 and 25 was used. Three neural network models were developed. Two of the models (identifying the top 33% and the lowest 33% groups, respectively) were able to reach 100% correct identification of all students in each of the two groups. The third model (identifying low, mid and high performance levels) reached precisions from 87% to 100% for the three groups. Analyses also explored the predicted outcomes at an individual level, and their correlations with the observed results, as a continuous variable for the whole group of students. Results demonstrate the greater accuracy of the ANN compared to traditional methods such as discriminant analyses. In addition, the ANN provided information on those predictors that best explained the different levels of expected performance. Thus, results have allowed the identification of the specific influence of each pattern of variables on different levels of academic performance, providing a better understanding of the variables with the greatest impact on individual learning processes, and of those factors that best explain these processes for different academic levels.

Keywords: Predictive systems; Academic performance; Artificial neural networks

Corresponding author: Mariel F. Musso, Katholieke Universiteit Leuven / Universidad Argentina de La Empresa, mariel.musso@hotmail.com

<http://dx.doi.org/10.14786/flr.v1i1.13>



1. Introduction

Many studies have explored the contribution to the explanation of academic performance with the use of various different variables and from diverse theoretical perspectives (e. g. Bekele & McPherson, 2011; Fenollar, Roman, & Cuestas, 2007; Kuncel, Hezlett, & Ones, 2004; Miñano, Gilar, & Castejón, 2008). Many factors have been identified as having important implications not only for the study of learning processes, but also as tools for improving of curriculum designs, tutorial systems, and students' academic results (Miñano et. al., 2008; Musso & Cascallar, 2009a; Zeegers, 2004). From this previous body of research, it has become apparent that the accurate prediction of student performance could have many useful applications for positive outcomes of the learning process and lead to advances in learning theory. For example, it could be helpful to identify students at risk of low academic achievement (Musso & Cascallar, 2009a; Ramaswami & Bhaskaran, 2010). This prediction could serve as an early warning of future low academic performance and guide interventions that could prove beneficial for such students. Similarly, being able to understand the role of different intervening variables that influence performance for all and for each category of performance level, would be a significant contribution to improve the approach to teaching and better understand learning processes. Many previous studies have focused on the prediction of academic performance (e.g., Hailikari, Nevgi, & Komulainen, 2008; Krumm, Ziegler, & Buehner, 2008; Turner, Chandler, & Heffer, 2009).

Many of the studies about academic performance have considered Grade Point Average (GPA) as the best summary of student learning, not only because of its strong prediction of performance for other levels of education (e. g. Kuncel et al., 2004, 2005), but also for other life outcomes as salary (Roth & Clarke, 1998), and job performance (Roth, Be Vier, Switzer, & Schippman, 1996).

The prediction of academic performance has been carried out with different methodological approaches. The first and most common approach found in the educational literature, has to do with the use of traditional statistical methods, such as discriminant analysis and multiple linear regressions (Braten & Stromso, 2006; Vandamme, Meskens & Superby, 2007). A second approach can be found in various studies which have used Structural Equation Modelling (SEM) to compare theoretical models to data sets and/or to test different models of academic performance (Fenollar et al., 2007; Miñano et al., 2008; Ruban & McCoach, 2005). These traditional approaches – that are tools widely used to predict GPA, to orient selection, placement, and/or classification of the academic process –failed to consistently show the capacity to reach accurate predictions or classifications in comparison with artificial intelligence computing methods (Everson, Chance, & Lykins, 1994; Kyndt, Musso, Cascallar, & Dochy, 2012, submitted; Lykins & Chance, 1992; Maucieri, 2003; Weiss & Kulikowski, 1991). Therefore, a third approach to the “prediction of academic performance” that we can find in recent literature involves machine learning techniques, such as methods using Artificial Neural Networks (ANN). This method has been used and proven useful in several other fields, such as business, engineering, meteorology, and economics. It is considered an important method to classify potential outcomes and is well regarded as an excellent pattern-recognizer (Detienne, Detienne, & Joshi, 2003; Neal & Wurst, 2001; White & Racine, 2001).

Recent work in the field of computer sciences has started to apply this methodology to large data banks of nation-wide educational outcomes (Abu Naser, 2012; Croy, Barnes, & Stamper, 2008; Fong, Si, & Biuk-Aghai, 2009; Kanakana, & Olanrewaju, 2011; Maucieri, 2003; Mukta & Usha, 2009; Pinninghoff Junemann, Salcedo Lagos, & Contreras Arriagada, 2007; Ramaswami & Bhaskaran, 2010; Zambrano Matamala, Rojas Díaz, Carvajal Cuello, & Acuña Leiva, 2011; Walczak, 1994). This methodology has also recently been used with various applications in educational measurement, in conjunction with other theoretical models of different constructs such as self-regulation of learning (Cascallar, Boekaerts & Costigan, 2006; Everson et al., 1994; Gorr, 1994; Hardgrave, Wilson, & Walstrom, 1994), reading readiness (Musso & Cascallar, 2009a); and performance in mathematics (Musso & Cascallar, 2009b; Musso, Kyndt, Cascallar, & Dochy, 2012). The application of predictive systems, with the emergence of new methodologies and technologies, have made it possible to assess a wide range of data and student performances in order to evaluate their current and future performance without the need for traditional testing (Boekaerts & Cascallar, 2006; Cascallar et al., 2006). This methodological approach using ANN can lead to the possible implementation of continuous assessment in the context of intelligent classrooms (Birenbaum et al., 2006).



Existing databases together with the constant monitoring of student performance could provide a continuous evaluation in real time of the students' progress.

The interrelationship between many of the variables participating in the complex and multi-faceted problem of academic performance are not clearly understood, and they are often related in nonlinear ways. ANN have demonstrated to be a very effective approach to address situations with these characteristics and to be able to classify and predict outcomes under those conditions with a high level of accuracy, especially when large data sets are available. This approach also allows the researcher to consider a large number of variables simultaneously and make use of their interrelationships without the usual parametric constraints. These advantages would allow researchers in the learning sciences to better understand the complex patterns of interactions between the variables at different levels of academic performance, not just for the prediction of performance but also to understand the participating factors that could be related to these outcomes. Several previous studies using ANN have addressed the classification of outcomes into different levels of performance, for different academic purposes: a) diagnostic purposes in order to identify those students most in need of support at the beginning of their primary school, regarding their readiness for learning to read (Musso & Cascallar, 2009a), and b) identifying students with low expected writing performance at the vocational secondary school level in order to provide support prior to their first year, and thus avoiding possible failure (Boekaerts & Cascallar, 2011). In these and other possible applications, the early detection of future low performance, and more targeted interventions, would decrease the negative experience of failure, and it would provide an important diagnostic tool for effective interventions. This approach would improve the chances of achieving successful outcomes, particularly for students identified as being "at-risk". Detecting and understanding the most significant variables that are the best indicators of the future low performers would be an important tool for management of school resources and planning remediation programs at all levels of an educational system.

Similarly, knowing the best indicators of the future high performers, would allow first of all the understanding of many of the factors leading to these positive outcomes. It would also allow an accurate selection of those students who could be assigned to advance programs, fellowships and/or be the object of talent searches. The accurate placement of students in different courses or programs according to how they are expected to perform would prevent possible failure, as well as providing the opportunity to offer challenging tasks for students expected to be among the high performers. In addition, a better understanding of the interrelationships between the variables leading to different levels of performance, would allow the fine-tuning of instructional approaches to the individual and/or group needs using the information provided by an ANN approach.

Some authors have shown that traditional statistical methods do not always yield accurate predictions and/or classifications (Bansal, Kauffman & Weitz, 1993; Everson, 1995; Duliba, 1991). Preliminary research using ANN for prediction, selection, and classification purposes suggests that this method may improve the validity and accuracy of the classifications, as well as increase the predictive validity of educational outcomes (Everson et al., 1994; Hardgrave et al., 1994; Perkins, Gupta, Tammana, 1995; Weiss & Kulikowski, 1991).

This paper explores this new methodological approach using a large amount of data collected from the students (including both cognitive and non-cognitive measures) in order to design predictive models using artificial neural networks (ANN). The ANN models in this research study can identify those predictors that could best explain different levels of academic performance in three different performance groups which cover all the range of performances, as well as making accurate classifications of the expected level of performance for each subject. Data about individual differences in basic cognitive variables were collected, since they are strongly related to the student's achievement (Colom, Escorail, Chin Shih, & Privado, 2007; Grimley & Banner, 2008). Although it has been argued that considering students' cognitive ability can lead to a relatively strong prediction of academic performance (Colom et al., 2007), this prediction could be strengthened by including background and non-cognitive predictors. As Chamorro-Premuzic & Arteche (2008) discuss, combining both cognitive ability and non-cognitive measures can provide a broader understanding of an individual's likelihood to succeed in academic settings, with models that predict such



performance at least one academic year in advance of the actual measure being obtained (grade-point average, GPA). In addition, discriminant analyses (DA) was used to analyse the same data in order to compare the predictive classificatory power of both methodologies. To better understand the rationale for this research, it is useful to review some of the main constructs included as predictors in this study, and to explain the quite novel methodology introduced from the family of predictive systems, that is, the machine learning modelling technique of Artificial Neural Networks (ANN).

2. Theoretical considerations

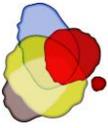
2.1 Working memory and academic performance

Intelligence and the g-factor are the most frequently studied factors in relation to academic achievement and the prediction of performance (Miñano et al., 2012). There is a large body of research that shows a strong positive correlation between g and educational success (e.g., Kuncel, Hezlett, & Ones, 2001; Linn & Hastings, 1984). The g-factor is defined, in part, as an ability to acquire new knowledge (e.g., Cattell, 1971; Schmidt, 2002; Snyderman & Rothman, 1987). Although the g-factor is not the same construct as Working Memory (WM), several studies have demonstrated a high correlation between these measures (Heitz et al., 2006; Unsworth, Heitz, Schrock, & Engle, 2005). Following the early study of Daneman and Carpenter (1980) on individual differences in working memory capacity (WMC) and reading comprehension, further research has shown the importance of WMC as a domain-general construct (Conway, Cowan, Bunting, Therriault, & Minkoff, 2002; Conway & Engle, 1996; Engle & Kane, 2004; Feldman Barrett, Tugade, & Engle, 2004; Kane et al., 2004), including the prediction of average scores over several academic areas (Colom et al., 2007).

Similarly, a large body of literature shows WMC as a very important construct in several areas and several studies have shown its importance in a wide range of complex cognitive behaviours such as comprehension (e.g., Daneman & Carpenter, 1980), reasoning (e.g., Kyllonen & Christal, 1990), problem solving (Welsh, Satterlee- Cartmell, & Stine, 1999) and complex learning (Kyllonen & Stephens, 1990; Kyndt, Cascallar, & Dochy, 2012; St Clair-Thompson & Gathercole, 2006). WMC is an important predictive variable of intellectual ability and academic performance, consistent over time (e.g. Engle, 2002; Musso & Cascallar, 2009a; Passolunghi & Pazzaglia, 2004; Pickering, 2006). Working memory is a paradigmatic form of cognitive control that explains how this cognitive control occurs, and which involves the active maintenance and executive processing of information available to the cognitive system, combining the ability to both maintain and effectively process information with minimal loss (Jarrold & Towse, 2006). It is crucial for the processing of information within the cognitive system, it has a limited capacity and it differs between individuals (Conway et al., 2005). The literature seems to indicate two fundamental approaches according to the interpretation of working memory and executive control. Traditional perspectives represent working memory and executive control as separate modules (e.g., Baddeley, 1986). The perspective taken in this research coincides with another view that understands working memory and executive control as constituting two sides of the same phenomenon, an emergent property from the neuro-cognitive architecture (Anderson, 1983, 1993, 2002, 2007; Anderson et al., 2004; Hazy; Frank & O'Reilly, 2006).

2.2 Attention and academic performance

Attention as a cognitive construct has been studied from different theoretical and methodological approaches (e.g., Posner & Rothbart, 1998; Redick & Engle, 2006; Rueda, Posner, & Rothbart, 2004). It is evident that our cognitive system is constantly receiving a variety of inputs from the environment. All these inputs are competing for the limited resources of the cognitive system, and requiring our "attention".



However, because human cognitive capacities are limited in their ability to process information simultaneously (Gazzaniga, Ivry, & Mangun, 2002), it is the shifting of the processing capacity and selection of stimuli to attend to, which constitute the basic aspects of our attentional system (Redick & Engle, 2006). This shifting and selection of incoming information is the function of the attentional system, which allows us to redirect our attention to the relevant aspects of the environmental information for the task or goals at hand. This study adopts the framework of Posner and Petersen (1990) who described three different and semi-independent attentional networks: orientation, alertness and executive attention. The orienting network allows the selection of information from sensory input, the alerting network refers to a system that achieves and maintains an alert state, and executive attention or executive control is responsible for resolving conflict among responses (Fan, McCandliss, Summer, Raz, & Posner, 2002). The efficiency of these three attentional networks can be quantified by reaction time measures (Fan et al., 2002). Redick and Engle (2006) and Unsworth et al. (2005) have found that individual differences in working memory capacity are related to those in attentional control, thus establishing that the executive control mechanism is closely related to working memory capacity.

Several studies have shown the importance of attention as a predictor of general academic performance (Gsanger, Homack, Siekierski, & Riccio, 2002; Kyndt et al., 2012, submitted; Riccio, Lee, Romine, Cash and Davis, 2002), reading (Landerl, 2010; Lovett, 1979), mathematical performance (Fernandez-Castillo & Gutiérrez-Rojas, 2009; Fletcher, 2005; Musso et al., 2012), and written expression (Reid, 2006). The research on learning disorders has found that attentional problems are negatively associated to academic achievement (Jimmerson, Dubrow, Adam, Gunnar, & Bozoky 2006).

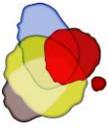
2.3 Learning strategies and academic performance

The estimated level of contribution of basic cognitive processes to the determination of academic achievement has shown considerable variation, which ranges from a moderate to a medium-high effect (Castejón & Navas, 1992; Navas, Sampascual, & Santed, 2003). Consequently, the studies focusing on the prediction of academic performance have increasingly included the so-called non-cognitive variables such as motivation, attributions, self-concept, effort, goal orientation, etc. (e.g., Fenollar et al., 2007; Pintrich, 2000). Learning strategies (LS) have been defined as student's actual behaviours, in a specific context, to engage in a task (Biggs, 1987). Other researchers describe LS as any thoughts or behaviours that help the students to acquire new information and integrate this new information with their existing knowledge (Weinstein & Mayer, 1986; Weinstein, Palmer, & Schulte, 1987; Weinstein, Schulte & Cascallar, 1982). LS also help students retrieve stored information. Examples of LS include summarizing, paraphrasing, imaging, creating analogies, note-taking, and outlining (Weinstein et al., 1987).

Previous research has provided support for the mediating role of learning strategies (Dupeyrat & Marine, 2005; Fenollar et al., 2007; Simons, Dewitte, & Lens, 2004). Fenollar et al. (2007) have compared a theoretical model, where achievement goals and self-efficacy were hypothesised to have direct effects on academic performance, to a mediating model where such effects were mediated through study strategies. Results from the study showed that achievement goals and self-efficacy have no direct effects on performance, and they suggest that the mediating model provides a better fit to the data (Fenollar et al., 2007).

2.4 Artificial neural networks and performance

Conceptually, a neural network is a computational structure consisting of several highly interconnected computational elements, known as neurons, perceptrons, or nodes. Each "neuron" or unit carries out a very simple operation on its inputs and transfers the output to a subsequent node or nodes in the network topology (Specht, 1991). Neural networks exhibit polymorphism in structure and parallelism in computation (Mavrouniotis & Chang, 1992), and it can be represented as a highly interconnected structure of processing elements with parallel computation capabilities (Grossberg, 1980, 1982; Rumelhart, Hinton, &



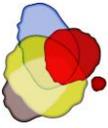
Williams, 1986; Rumelhart, McClelland, & the PDP research group, 1986). In general, an ANN consists of an input layer (which can be considered the independent variables), one or more hidden layers, and an output layer that is comparable to a categorical dependent variable (Cascallar et al., 2006; Garson, 1998). All ANN process data through multiple processing entities which learn and adapt according to patterns of inputs presented to them, by constructing a unique mathematical relationship for a given pattern of input data sets on the basis of the match of the explanatory variables to the outcomes for each case (Marshall & English, 2000). Thus, neural networks construct a mathematical relationship by “learning” the patterns of all inputs from each of the individual cases used in training the network, while more traditional approaches assume a particular form of relationship between explanatory and outcome variables and then use a variety of fitting procedures to adjust the values of the parameters in the model.

During the training phase, ANNs generate a predicted outcome for each case, and when this prediction is incorrect the network makes adjustments to the weights of the mathematical relationships among the predictors and with the expected outcome, weights that are represented in the hidden layers of the network. The predicted output is a continuous variable with a specific value for each case (or subject) which includes information on the probability of belonging to each of the categorical classifications requested by the developer of the ANN. According to this architecture, the ANN finally recognizes patterns and classifies the cases presented into the requested outcome categories, depending on the target question, and given the individual probability values for each case. This information is generated by the network through many iterations, gradually changing and adjusting the weights for all the interrelationships between the units after each incorrect prediction. During this training process, the network becomes increasingly accurate in replicating the known outcomes from the test cases. The neural network continues to improve its predictions until one or more of the pre-determined stopping criteria have been met. These stopping criteria can be, for example, a minimum level of accuracy, learning rate, persistency, number of iterations, amount of time, etc.

Once trained, the network is tested with the remaining cases in the dataset, which is considered a form of validation of the network (testing phase), by observing how the weights in the model, now fixed to those obtained in the training phase, predict classes of outcomes in a new set of data of which outcomes are known to the experimenter but not to the ANN system. Afterwards it can also be applied to predict future cases where the outcome is still unknown (Cascallar et al., 2006). In addition, with complementary techniques in predictive stream analysis, the neural network approach allows us to determine the predictive power of each of the variables involved in the study, providing information about the importance of each input variable (Cascallar et al., 2006; Garson, 1998).

Predictive stream analyses (Cascallar & Musso, 2008), based in this case on neural network (ANN) models, have several strengths: (a) because these are machine learning algorithms, the assumptions required for traditional statistical predictive models (e.g., ordinary least squares regression) are not necessary. As such, this technique is able to model nonlinear and complex relationships among variables. ANN aim to maximize classification accuracy and work through the data in an interactive process until maximum accuracy is achieved, automatically modelling all interactions among variables; (b) ANNs are robust, general function estimators. They usually perform prediction tasks at least as well as other techniques and most often perform significantly better (Marquez, Hill, Worthley, & Remus, 1991); (c) ANN can handle data of all levels of measurement, continuous or categorical, as inputs and outputs. Because of the speed of microprocessors in even basic computers, ANNs are more accessible today than when they were originally developed. Current research has shown that neural network analysis substantially improves the validity of the classifications and increases the accuracy and predictive validity of the models, in education and other fields (Kyndt et al., 2012, submitted; Musso & Cascallar, 2009b; Perkins et al., 1995).

The ANN learns by examining individual training cases (subjects/students), then generating a prediction for each student, and making adjustments to the weights whenever it makes an incorrect prediction. Information is passed back through the network in iterations, gradually changing the weights. As training progresses, the network becomes increasingly accurate in replicating the known outcomes. This process is repeated many times, and the network continues to improve its predictions until one or more of the stopping criteria have been met. A minimum level of accuracy can be set as the stopping criterion, although



additional stopping criteria may be used as well (e.g., number of iterations, amount of processing time). Once trained, the network can be applied, with its structure and parameters, to future cases (validation or holdout sample) for further validation studies and programme implementation (Lippman, 1987). As long as the basic assumptions of the population of persons or events that the ANN used for training is constant or varies slightly and/or gradually, it can adapt and improve its pattern recognition algorithms the more data it is exposed to in the implementations.

The class of ANN models used in this research can be compared with the more traditional discriminant analysis approach. Both of these methods derive classification rules from samples of classified objects based on known predictors. This general approach is called ‘supervised learning’ since the outcomes are known and relationships are modelled or ‘supervised’ according to these outcomes (Kohavi & Provost, 1998). But, there are significant differences in the algorithms and procedures for both analyses, such as the fact that while discriminant analysis assumes linear relationships, neural network analysis does not. In terms of comparisons with another common statistical method used in educational research, linear regression, it is important to note that although neural networks can address some of the same research issues as regression it is inherently a different mathematical approach (Detienne et al., 2003). There is another family of predictive systems which are “unsupervised” (e.g., Kohonen networks), in which the patterns presented to the network are not associated with specific outcomes; it is the neural network itself that derives the commonalities between the predictors, grouping cases into classes on the basis of these similarities. Thus, these analyses can be used to explore the data from a different perspective and learn the grouping of cases based on these predictor commonalities instead of being focused on predictions or individual outcomes (Cascallar et al., 2006; Kyndt et al., 2012, submitted).

Neural networks excel in the classification and prediction of outcomes; especially when large data sets are available that are related in nonlinear ways, and where the intercorrelation between variables is not clearly understood. These properties of ANNs clearly make them particularly suitable for social science data where they can simultaneously consider all variables in a study (Garson, 1998). Moreover, the assumptions of normality, linearity and completeness that are made by methods such as multiple linear regression (Kent, 2009), and that are often very difficult to establish for social science data, are not made in neural network analysis. Neural networks can work with noisy, incomplete, overlapping, highly nonlinear and non-continuous data because the processing is spread over a large number of processing entities (Garson, 1998, Kent, 2009). In this regard it can be said that neural networks are robust and have wide non-parametric application. There is also evidence that neural models are robust in the statistical sense, and also robust when faced with a small number of data points (Garson, 1998).

Very few studies within the educational literature have used neural network analysis or any other type of predictive system (e.g., Cascallar et al., 2006; Cascallar & Musso, 2008; Musso & Cascallar, 2009a; Pinninghoff Junemann et al., 2007; Wilson & Hardgrave, 1995).

2.5 ANN processing and measures to evaluate the neural network system performance

In order to evaluate the performance of the neural network system, there are a number of measures used which provide a means of determining the quality of the solutions offered by the various network models tried. The traditional measures include the determination of actual numbers and rates for True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) outcomes, as products of the ANN analysis. In addition, certain summative evaluative algorithms have been developed in this field of work, to assess overall quality of the predictive system.

These overall measures are: Recall, which represents the proportion of correctly identified targets, out of all targets presented in the set, and is represented as: $\text{Recall} = \text{TP}/(\text{TP} + \text{FN})$; and Precision which represents the proportion of correctly identified targets, out of all identified targets by the system, and is represented as: $\text{Precision} = \text{TP}/(\text{TP} + \text{FP})$. Two other measures, derived from signal-detection theory (ROC analysis), have also been used to report the characteristics of the detection sensitivity of the system. One of them is Sensitivity (similar to Recall: the proportion of correctly identified targets, out of all targets



presented in the set), and which is expressed as $\text{Sensitivity} = \text{TP}/(\text{TP} + \text{FN})$. The other is Specificity, defined as the proportion of correctly rejected targets from all the targets that should have been rejected by the system, and which is expressed as $\text{Specificity} = \text{TN}/(\text{TN} + \text{FP})$. All the traditional measures are typically represented in what is called a “confusion matrix” representing all four outcomes.

In addition, the evaluation of ANN performance is also carried out with another summative measure, which is used to account for the somewhat complementary relationship between Precision and Recall. This measure is defined as F_1 , and is defined as $F_1 = (2 * \text{Precision} * \text{Recall})/(\text{Precision} + \text{Recall})$. Such a definitional expression of F_1 assumes equal weights for Precision and Recall. This assumption can be modified to favour either Precision or Recall, according to the utility and cost/benefit ratio of outcomes favouring either Precision or Recall for any given predictive circumstance.

2.6 Objectives and research questions

The objective of this study is to identify patterns of variables that will allow a correct predictive classification of three levels of General Academic Performance (GAP) into: Low, Middle and High GAP, measured by the grade-point-average (GPA). This was achieved by taking into consideration basic cognitive processes (working memory capacity; alerting, orienting and executive attention), learning strategies, and family-social background factors. The idea behind this paper is to explore new approaches to obtain predictive classifications of learning outcomes, without the use of one specific test, using a large number of variables (cognitive and non-cognitive) that could better capture the true complex composite of influences participating in the actual observed outcomes from individual students. In addition, it is another objective of the research to explore the differences in the patterns predicting each level of performance (low, middle and high performance) to inform future research into the causal factors generating and participating in those sets of identified variables and that could explain different levels of performance using artificial neural networks. Of course, previous academic performance could have been taken into account to facilitate the predictive classification, but this was purposely avoided for two reasons: as a proof-of-concept that other variables are sufficient to predict academic performance, and to highlight more clearly the weight that each of these other variables has in the determination of a student’s academic performance.

In order to explore the differences in the patterns predicting each level of performance, three artificial neural network (ANN) models were developed. Two of them to predict the students who would be in each of the extreme performance levels (low 33% and high 33% of GPA) in order to analyse the differences between the patterns of variables having the most predictive weight for each group, and thus providing information on the potentially different processes involved in those low and high performance outcomes. A third ANN was developed, capable of accurately producing a predictive classification for the three levels of performance simultaneously (low 33%, middle 33%, and high 33%). This final ANN model was capable of finding the common patterns that could predict simultaneously all performance groups. The relative importance of the predictors for each network was also analysed. The predictive capability of each ANN was systematically improved by modifying the parameters that determine the rate of learning, the persistence, momentum, and stopping criteria, and the type of functions used for weight adjustments. Precision, sensitivity, specificity and accuracy of the three networks were obtained. In addition, the correlation between the individual prediction for each student and the actual observed GPA was established, and proved to be very high.

The main research questions of this study are: How accurately can different levels of academic performance in higher education be predicted by working memory capacity, attentional networks, learning strategies and background variables when used as inputs in a neural network model? What is the relative importance of the predictor variables and the observed differences for each performance level category?



3. Method

3.1 Participants

The total sample included 864 university students, of both genders (male 45.4%; female 54.6%), ages between 18 and 25 ($M_{age} = 20.38$, $SD = 3.78$), recently enrolled in the first year in several different disciplines (psychology, engineering, medicine, law, social communication, business and marketing), in three private universities in Argentina, during the 2009-2011 academic years. In all, 67.8% of the sample was 17 to 20 years old, 24.7% was 21-25 years old, and 7.5% was older than 25 years. The students in the sample came from private religious secondary schools (48.5%), private non-religious schools (19%), private bilingual schools (15.4%), public secondary schools (15%), and 2.1% from international community schools. All student data (predictors) was collected at the beginning of the corresponding academic year, and the dependent variable (GPA) was collected at the end of the same academic year. An 80% math accuracy criterion was imposed for all participants in the Automated Operation Span (Unsworth et al., 2005). Therefore, they were encouraged to keep their math accuracy at or above 80% at all times (to insure that the interfering task was actually being performed). As a consequence of this criterion, 78 participants were excluded from the analyses. The final sample consisted of 786 students.

3.2 Instruments

3.2.1 Attention Network Test (ANT) (Fan et al., 2002)

This computerized task provides a measure for each of the three anatomically defined attentional networks: alerting, orienting, and executive. The ANT is a combination of the cued reaction time (Posner, 1980) and the flanker test (Eriksen & Eriksen, 1974). The participant saw an arrow on the screen that, on some trials, was flanked by two arrows to the left and two arrows to the right. Participants were asked to determine when the central arrow points left or right, by two mouse buttons (left- right). They were instructed to focus on a centrally located fixation cross throughout the task, and to respond as quickly and accurately as possible. During the practice trials, but not during the experimental trials, subjects received feedback from the computer on their speed and accuracy. The practice trials took approximately 2 minutes and each of the three experimental blocks took approximately 5 minutes. The whole experiment took about twenty minutes. The measure for (general) attention is the average response time regardless of the cues or flankers. To analyse the effect of the three attentional networks, a set of cognitive subtractions described by Fan et al. (2002) were used. The efficiency of the three attentional networks is assessed by measuring how response times are influenced by alerting cues, spatial cues, and flankers (Fan et al., 2002). The alerting effect was calculated by subtracting the mean response time of the double-cue conditions from the mean response time of the no-cue conditions. For the orienting effect, the mean response time of the spatial cue conditions (up and down) were subtracted from the mean response time of the center cue condition. Finally, the effect of the executive control (conflict effect) was calculated by subtracting the mean response time of all congruent flanking conditions, summed across cue types, from the mean response time of incongruent flanking conditions (Fan et al. 2002). The test-retest reliability of the general response times (in this study used as a measurement of general attention), calculated by Fan et al. (2002) equaled .87. The test-retest reliability of the subtractions is less good. The executive control is the most reliable ($r=.77$), followed by the orienting network ($r=.61$). The alerting network showed to be the least reliable ($r=.52$) (Fan et al. 2002).

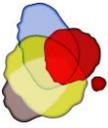


3.2.2 Automated Operation Span (Unsworth et al., 2005)

This is a computer-administered version of the Ospan instrument (Unsworth et al., 2005) that measures working memory capacity. The responses were collected via click of a mouse button. First, participants receive practice and secondly, the participants perform the actual experiment. The practice sessions are further broken down into three sections. The first practice is a simple letter span task. They see letters appear on the screen one at a time. In all experimental conditions, letters remain on-screen for 800 milliseconds (ms). Then, participants must recall these letters in the same order they saw them from a 4 x 3 matrix of letters (F, H, J, K, L, N, P, Q, R, S, T, and Y) presented to them. Recall consists of clicking the box next to the appropriate letters; the recall phase is untimed. After each recall, the computer provides feedback about the number of letters correctly recalled. Next, participants practice the math portion of the experiment. Participants first see a math operation (e.g. $(1*2) + 1 = ?$). Once the participant knows the answer they click the mouse to advance to the next screen. Participants then see a number (e.g. “3”) and are required to click if the number is the correct solution by clicking on “True” or “False.” After each operation participants are given feedback. The math practice serves to familiarize participants with the math portion of the experiment, as well as to calculate how long it takes a given person to solve the math problems, establishing an individual baseline. Thus, it attempts to account for individual differences in the time it takes to solve math problems. This is then used as an individualized time limit for the math portion of the experimental session. The final practice session has participants perform both the letter recall and math portions together, just as they will do in the experimental block. The participants first are presented with a math operation, and after they click the mouse button indicating that they have solved it, they see the letter to be recalled. If the participants take more time to solve the math operations than their average time plus 2.5 SD, the program automatically moves on and counts that trial as an error. This serves to prevent participants from rehearsing the letters when they should be solving the operations. Participants complete three practice trials, each of set size 2. After the participant completes all of the practice sessions, the program moves them on to the real trials. The real trials consist of 3 sets of each set-size, with the set-sizes ranging from 3 to 7 letters. This makes for a total of 75 letters and 75 math problems. Subjects are instructed to keep their math accuracy at or above 85% at all times. During recall, a percentage in red is presented in the upper right-hand corner. Subjects are instructed to keep a careful watch on the percentage in order to keep it above 85%. This study reports the Absolute Ospan score (the sum of all perfectly recalled sets) that is interpreted as the measure of overall working memory capacity, and one Reaction Time score (operations). The task takes approximately 20–25 minutes to complete (Unsworth et al., 2005). This measure of working memory capacity has a high correlation with other measures of working memory and general intelligence, as Ospan and Raven Progressive Matrices. In addition, AOSPAN has a good test-retest reliability ($r = .83$) and an adequate internal consistency ($\alpha = .78$) (Unsworth et al., 2005).

3.2.3 Learning Strategies Questionnaire (LASSI; Weinstein et al., 1987; Weinstein & Palmer, 2002; Weinstein et al., 1982).

The original version is a 77-item questionnaire with 10 scales that assesses the students' awareness about, and use of, learning and study strategies related to skill, will, and self-regulation components of strategic learning. These scales and their corresponding internal consistency coefficients reported in the Users' Manual (Weinstein & Palmer, 2002), are as follows: Attitude Scale ($\alpha = .77$), Motivation Scale ($\alpha = .84$), Time Management Scale ($\alpha = .85$), Anxiety Scale ($\alpha = .87$), Concentration Scale ($\alpha = .86$), Information Processing Scale ($\alpha = .84$), Selecting Main Ideas scale ($\alpha = .89$), Study Aids Scale ($\alpha = .73$), Self-Testing Scale ($\alpha = .84$), and Test Strategies Scale ($\alpha = .80$). The present study used a Spanish-version (Strucchi, 1991), which was slightly modified in some semantic and grammatical aspects for the local sample. The exploratory factor analysis determined a matrix with five factors that explained 37.52% of the variance. Factor 1 related to “cognitive resources/cognitive processing” ($\alpha = .871$; 13 items; $R^2 = 18.03\%$); Factor 2, related to “time management” ($\alpha = .807$; 10 items; $R^2 = 8.404\%$); Factor 3, dealing with “processing of information and generalization” ($\alpha = .783$; 8 items; $R^2 = 4.567\%$); Factor 4 which is related to “anxiety management” ($\alpha = .60$; 5 items; $R^2 = 3.431\%$); and Factor 5, which involves the construct of “study



techniques and use of help” ($\alpha = .728$; 7 items; $R^2 = 2.685\%$). Students gave responses on a Likert-type scale, from 1 (never) to 5 (always).

3.2.4 Background information

Basic background information of each student used in the analyses was: gender, highest level of education of mother and father (not completed primary school- primary school- secondary school- graduated university- post-graduate), occupation of parents, and secondary school from which the student graduated (public - private religious school - private non-religious school - bilingual school - foreign community)

3.2.5 Academic performance

Academic performance was measured by the Grade Point Average (GPA) of all courses (different subjects depending on the discipline) at the end of each of the academic years. All course grades which are used by the universities to calculate the overall GPA are obtained using university-wide criteria for the interpretation and assignment of final scores in each course, from which the GPA was calculated. The GPA information was collected from official records at the end of the first academic year for each student, at each of the participating universities, and they all are in a scale from 0 to 10 (with 10 indicating best performance).

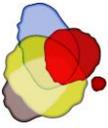
3.3 Analyses procedure

The ANN model used was a backpropagation multilayer perceptron neural network, that is, a multilayer network composed of nonlinear units, which computes its activation level by summing all the weighted activations it receives and which then transforms its activation into a response via a nonlinear transfer function, which establishes a relationship between the inputs and the weights they are assigned. During the training phase, these systems evaluate the effect of the weight patterns on the precision of their classification of outputs, and then, through backpropagation, they adjust those weights in a recursive fashion until they maximize the precision of the resulting classifications.

ANN parameters and variable groupings, as well as all other network architecture parameters, were adjusted to maximize predictive precision and total accuracy. Confusion matrices have been determined for each ANN, as well as ROC analyses for the evaluation of sensitivity and specificity parameters. Parameters such as learning rate (the rate at which the ANN “learns” by controlling the size of weight and bias changes during learning), momentum (adds a fraction of the previous weight update to the current one, and is used to prevent the system from converging to a local minimum), number of hidden layers, stopping rules (when the network should stop “learning” to avoid over-fitting the current sample), activation functions (which define the output of a node given an input or set of inputs to that node or unit), and number of nodes were specified and varied in the model construction phase in order to maximize the overall performance of the network model.

3.4 Architecture of the neural networks

According to the objectives of this research, three different neural networks (ANN) were developed as predictive systems for the GPA of the students in this study. ANN₁ was developed to maximize the predictive classification of the lowest 33% of students, which would be scoring the lowest average GPA at the end of the academic year. ANN₂ was developed to maximize the predictive classification of the highest 33% of students, which would be scoring the highest GPA. ANN₃ was developed to predict the classification of students into the three levels of expected GPA at the same time. The data set was partitioned into a training set and a testing set for each ANN, and for each network, training and testing samples were chosen at random by the software, from the available set of cases. One suggested criterion is that the number of



training inputs (cases) should be at least 10 times the number of input and middle layer neurons in the network (Garson, 1998). Similarly, it is suggested that about 2/3 (or 3/4) of the cases in the available data set be used for the training phase in order to include a set of cases representing most of the patterns expected to be present in the data (patterns represented by the vector for each case). The remaining 1/3 or 1/4 of the data is used for the testing phase of the network. The specific architecture of each of the three neural networks developed is as follows:

ANN₁ - (Maximizing the prediction for the Low 33% performance group): All cognitive variables, learning strategies, and background variables were introduced in the analysis. They were used for the development of the vector-matrix containing all predictor variables for each student. The resulting network contained all the input predictors, with a total of 18 input units (Reaction Time Operation, Reaction Time Math, Reaction Time Problem, Orienting Attention, Alerting Attention, Executive Control, Absolute Aospan, Processing of information/ Generalization, Study Techniques and use of help, Anxiety Management, Time Management, Cognitive resources/Cognitive processing, Gender, Mother's occupation, Father's occupation, Secondary school from which the student graduated, Highest level of education completed by father, and Highest level of education completed by mother). The model built contained one hidden layer, with 15 units. The output layer contained a dependent variable with two units (categories corresponding to “belongs to lowest 33%” or “belongs to highest 67 %”). In terms of the architecture of the network, a standardized method for the rescaling of the scale dependent variables was used. The hidden layer had a hyperbolic tangent activation function which is the most common activation function used for neural networks because of its greater numeric range (from -1 to 1) and the shape of its graph. The output layer utilized a softmax activation function that is useful predominantly in the output layer of a clustering system, converting a raw value into a posterior probability. The output layer used the cross-entropy error function in which the error signal associated with the output layer is directly proportional to the difference between the desired and actual output values. This function accelerates the backpropagation algorithm and it provides good overall network performance with relatively short stagnation periods (Nasr, Badr, & Joun, 2002). The training was carried out with the ‘online’ methodology (one case per cycle), with an initial learning rate of 0.4, and momentum equal to 0.9. The optimization algorithm was gradient descent (which takes steps proportional to the negative of the approximate gradient of the function at the current point), and the minimum relative change in training error was 0.0001.

ANN₂ - (Maximizing the prediction for the High 33% performance group): All cognitive, learning strategies, and background variables were introduced in the analysis. They were used for the development of the vector-matrix containing all predictor variables for each student. The resulting network contained all the input predictors, with a total of 18 units (Reaction Time Operation, Reaction Time Math, Reaction Time Problem, Orienting Attention, Alerting Attention, Executive Control, Absolute Aospan, Processing of information/Generalization, Study Techniques and use of help, Anxiety Management, Time Management, Cognitive resources/Cognitive processing, Gender, Mother's occupation, Father's occupation, Secondary school from which the student graduated, Highest level of education completed by father, and Highest level of education completed by mother). The model built contained one hidden layer, with nine units, and an output layer with two units (categories corresponding to “belongs to highest 33%” or “belongs to lowest 67%”). In terms of the architecture of the network, a standardized method for the rescaling of scale dependent variables was used. The hidden layer had a hyperbolic tangent activation function. The output layer utilized a softmax activation function. Cross-entropy was chosen as the error function. The dataset was partitioned into training set and testing set. The training was carried out with the ‘online’ methodology, with an initial learning rate of 0.5, and momentum equal to 0.7. The optimization algorithm was gradient descent, and the minimum relative change in training error was 0.0001.

ANN₃ - (Maximizing the simultaneous prediction for all the performance groups: Low 33% - Middle 33% - High 33%, simultaneously): All cognitive, learning strategies and background variables were introduced in the analysis. They were used for the development of the vector-matrix containing all predictor



variables for each student. The resulting network contained all the input predictors, with a total of 19 input units (Reaction Time Operation, Reaction Time Math, Reaction Time Problem, Orienting Attention, Alerting Attention, Executive Control, Absolute Aospan, Processing of information/ Generalization, Study Techniques and use of help, Anxiety Management, Time Management, Cognitive resources/Cognitive processing, Gender, Mother's occupation, Father's occupation, Secondary school, Highest level of education completed by father, and Highest level of education completed by mother, Ln of Attention Total RT). The model built contained one hidden layer, with 20 units, and one output layer with three units (categories corresponding to “belongs to low 33%”, “belongs to middle 33%” or “belongs to high 33%” of the performance groups). In terms of the architecture of the network, a standardized method for the rescaling of scale dependent variables was used. The hidden layer and the output layer both had a hyperbolic tangent activation functions. A standardized method for the rescaling of covariates was used. Sum of squares was chosen as error function. The dataset was partitioned into training set and testing set. The training was carried out with the ‘online’ methodology, with an initial learning rate of 0.4, and momentum equal to 0.8. The optimization algorithm was gradient descent, and the minimum relative change in training error was 0.0001.

The software used was SPSS v.19 – Neural Network Module, for the development and analysis of all predictive models in this study. Two development phases of the predictive system were carried out: training of the network and testing of the network developed. During the training phase several models were attempted, and several modifications of the neural network parameters were explored, such as: learning persistence, learning rate, momentum, and other criteria. These tests continued until achieving desired levels of classification, maximizing the benefits of the model chosen. In these analyses both precision and recall, as outcome measures of the network, were given equal weight. There was no need to trim the number of predictor inputs in the three models. The validation procedure used was the leave-one-out methodology.

3.5 Discriminant analyses

Discriminant Analyses (DA) were carried out using the same data and the same categories of GPA used in the Neural Networks Analyses. DA₁ was performed to discriminate between the students belonging to the lowest 33% of GPA and contrasting them against those not in that category. DA₂ was focused on identifying students in the highest 33% of academic performance versus those not in that group, and DA₃ was calculated to discriminate the students belonging to each one of the three levels of GPA performance. In order to give every variable the opportunity to contribute significantly to the prediction, a stepwise discriminant analysis was calculated for each category including all independent variables. In addition, we calculated three discriminant analyses, one for each category including the independent variables of the maximised neural networks of each category.

4. Results

4.1 Descriptive data

The final sample included 786 university students from several disciplines (Psychology, Engineering, Medicine, Law, Social Communication, Business and Marketing), in three private universities, during the 2009-2011 academic years.

Descriptive statistics of the cognitive variables and learning strategies are presented in Table 1 (cognitive variables) and Table 2 (learning strategies).

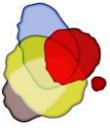


Table 1

Descriptive Statistics for Attentional Networks, General Reaction Time, Working Memory Capacity (Absolute Aospan) and Reaction Time Operation

	Alerting Attention	Orienting Attention	Executive Control	Ln of Attention Total RT	Absolute Aospan (Sum of perfectly recalled sets)	Ln RT Operation
N	786	786	786	786	786	786
Mean	34.40	44.01	102.54	6.20	27.88	7.01
SD	22.14	22.90	41.68	.11	14.83	.20
Skewness	.25	.24	3.31	.67	.25	.46
Kurtosis	1.96	5.01	26.14	.98	-.510	.45
Minimum	-78.00	-77.67	19.00	5.92	0	6.50
Maximum	123.83	213.83	558.00	6.74	68	7.75

Note: Ln of Attention Total RT: Logarithm of Attention Total Reaction Time (measure of Attention Network Test)

Ln RT Operation: Logarithm of Reaction Time Operation (measure of AOSPAN)

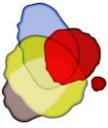
Table 2

Descriptive Statistics for Each Factor of Learning Strategies (LASSI)

	Cognitive resources/Cognitive processing	Time Management	Processing of information/Generalization	Anxiety Management	Study Techniques and use of help
N	756	756	756	756	756
Mean	-.02	.00	.01	.00	-.01
SD	1.09	1.12	1.11	1.15	1.14
Skewness	.24	.18	-.37	.35	-.67
Kurtosis	-.16	-.21	-.07	-.41	-.03
Minimum	-2.87	-2.86	-4.61	-2.53	-4.24
Maximum	3.85	3.30	2.56	3.57	2.22

4.2 Neural network analyses

ANN₁ was designed to predict the performance group corresponding to the lowest 33% of predicted GPA. It included 82.4 % of the participants (n = 632) in the training phase and 17.6% (n = 111) in the testing phase. After training, ANN₁- predicting the group with the low 33% of academic performance – was able to reach 100% correct identification of the students that belong to the target group (Lowest 33%) (see Figure 1). The precision of ANN₁ equalled 1 on a maximum of 1. The sensitivity of the network equalled 1, and the specificity (defined as the proportion of correctly rejected targets from all the targets that should have been rejected by the system) was equal to 1. The area under the curve equalled .877.



		Prediction of academic performance	
		33% Lowest (target group)	Others
Observed academic performance	33% Lowest (target group)	100%	0%
	Others	0%	100%

Figure 1. Testing Phase of the Neural Network Predicting the Lowest 33% of Academic Performance Scores.

In general, several tables (3-5) show the actual predictive weights of the variables that the ANNs used in the prediction of future academic performance for each of the groups (Low 33%, High 33% and the whole sample). The “Importance” column can be interpreted as the actual predictive weight of each variable, and the “Normalized Importance” column represents the percent of predictive weight for each variable (in each group’s analysis) with respect to the variable with the greatest predictive weight for the group in question, which is assigned a 100%. Table 6 summarizes the actual predictive weights of the variables, grouped by construct: Background variables (i.e., parents’ education, parents’ occupation, type of secondary school), Basic Cognitive variables (i.e., working memory capacity, attentional networks), Reaction time variables (i.e., operations, attentional), and Learning Strategies/Motivation variables (i.e., study techniques, time management, anxiety management). It allows an easier comparison of the sources of predictive weights by area between the various student groups and also for the total sample.

Table 3 shows the actual predictive weight of each input, and the normalised importance of the different variables for the ANN₁ predictive classification. These results indicate that the learning strategies regarding cognitive processes, reaction time (RT), and time management were the most important predictors. All reaction times are converted to natural logarithms (Ln) of the actual RT.

Table 3

Relative Importance of the Most Predictive Variables included in the Model for the Predictive Classification of the Lowest 33% of Scores in Academic Performance

Low 33% Group		
Independent Variable Importance		
Variables	Importance	Normalized Importance
Cognitive resources/Cognitive processing	0.092	100.00%
Ln Reaction Time Math	0.083	90.80%
Time Management	0.080	87.30%
Secondary school from which the student graduated	0.066	71.50%
Father's occupation	0.065	70.90%
Executive Control	0.062	67.60%
Mother's occupation	0.058	63.70%
Ln Reaction Time Problem	0.058	62.80%



Absolute Aospan (Sum of perfectly recalled sets)	0.055	60.50%
Anxiety Management	0.051	55.40%
Alerting Attention	0.050	54.40%
Ln Reaction Time Operation	0.048	52.40%
Orienting Attention	0.048	52.10%
Study Techniques and use of help	0.046	51.70%
Processing of information/ Generalization	0.043	46.50%
Gender	0.040	43.70%
Highest level of education completed by mother	0.030	32.60%
Highest level of education completed by father	0.025	27.10%

ANN₂ was designed to predict the performance group corresponding to the highest 33% predicted GPA. It included 77.9% of the students in the training phase ($n= 614$) and 22.1% in the testing phase ($n= 136$). After training, ANN₂ reached an accuracy of 100 % (see Figure 2). The precision of ANN₂ equalled 1 on a maximum of 1. The sensitivity of the network equalled 1, and the specificity amounted to 1. The area under the curve equalled .788.

		Prediction of academic performance	
		<i>33% Highest (target group)</i>	<i>Others</i>
Observed academic performance	<i>33% Highest (target group)</i>	100%	0%
	<i>Others</i>	0%	100%

Figure 2. Testing Phase of the Neural Network Predicting the Highest 33% of Academic Performance Scores.

The most important variables for the prediction of ANN₂ (High 33%) were reaction time, mother’s occupation, type of secondary school, father’s occupation and executive control (executive attention measure) (see Table 4).

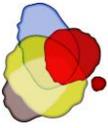


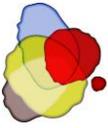
Table 4

Relative Importance of the Most Predictive Variables included in the Model for the Predictive Classification of the Highest 33% of Scores in Academic Performance

High 33% group		
Independent Variable Importance		
Variables	Importance	Normalized Importance
Ln of Reaction Time Operation	0.084	100.00%
Mother's occupation	0.081	97.10%
Secondary school from which the student graduated	0.081	96.10%
Father's occupation	0.076	90.10%
Executive Control	0.072	86.40%
Alerting Attention	0.062	73.90%
Processing of information/ Generalization	0.055	65.10%
Orienting Attention	0.054	64.10%
Study Techniques and use of help	0.053	62.30%
Highest level of education completed by father	0.051	60.70%
Ln of Reaction Time Math	0.049	58.50%
Anxiety Management	0.047	55.60%
Highest level of education completed by mother	0.044	52.80%
Absolute Aospan (Sum of perfectly recalled sets)	0.044	52.70%
Time Management	0.044	52.20%
Cognitive resources/Cognitive processing	0.037	44.70%
Ln of Reaction Time Problem	0.033	39.90%
Gender	0.033	39.60%

Both networks showed interesting differences in the pattern of relative normalized importance of those variables with the highest participation in the predictive model. For the low performing group in terms of general GPA (those predicted to be in the lowest 33% of scores), several learning strategies related to cognitive processes, reaction time (WMC and attentional networks functioning), and time management were most important in providing predictive weights for a correct classification. On the other hand, results from the predictive model for those students expected to be in the highest 33% of the general GPA scores, the top three predictors with the most significant participation were background variables involving mother's and father's occupation, type of secondary school, and overall reaction time of the cognitive and attentional processes.

ANN₃, which was designed to predict the three GPA performance groups simultaneously, used 82.8% of the students ($n=710$) for the training phase, and 17.2% ($n=122$) for the testing phase. After maximizing the training procedures, the accuracy in the testing phase reached 87.5% for the Lowest 33%, 100% for the Middle 33%, and 100% for the Highest 33% (see Figure 3). The precision of ANN₃ equalled .875 on a



maximum of 1. The sensitivity of the network equalled 1, and the specificity amounted to .50. The areas under the curve were .658 for the Low 33%, .583 for the Middle 33%, and .637 for the High 33%.

		Prediction of academic performance		
		33% Lowest	Middle 33%	33% Highest
Observed academic performance	Low 33%	87.5 %	10%	2.5%
	Middle 33%	0%	100%	0%
	High 33%	0%	0%	100%

Figure 3. Testing Phase of the Neural Network Predicting the Three Levels of Academic Performance Scores (Low 33%- Middle 33%- High 33%).

The most important variables for the prediction of ANN₃ were orienting attention, learning strategies related to the cognitive resources and information processing, time management, and executive control (executive attentional network) (see Table 5).

Table 5

Relative Importance of the Most Predictive Variables included in the Model for the Predictive Classification of the Three Levels of Academic Performance

All 3 Groups - GPA (Low 33% - Mid 33% - High 33%)		
Independent Variable Importance		
Variables	Importance	Normalized Importance
Orienting Attention	0.087	100.00%
Cognitive resources/Cognitive processing	0.076	86.86%
Time Management	0.074	84.92%
Executive Control	0.073	83.30%
Father's occupation	0.071	81.80%
Mother's occupation	0.070	79.91%
Ln of Attention Total Reaction Time	0.067	77.25%
Alerting Attention	0.067	76.63%
Ln of Reaction Time Math	0.061	70.14%
Processing of information/ Generalization	0.050	57.20%
Ln of Reaction Time Operation	0.043	49.64%
Study Techniques and use of help	0.041	46.55%
Ln of Reaction Time Problem	0.040	46.13%



Anxiety Management	0.038	43.89%
Gender	0.032	36.67%
Highest level of education completed by father	0.031	35.73%
Absolute Aospa (Sum of perfectly recalled sets)	0.031	35.09%
Highest level of education completed by mother	0.026	29.88%
Secondary school	0.024	27.29%

4.3 Maximizing the ANN models

All ANN models were developed so as to maximize the accuracy of the classification. The number of units in the hidden layers was determined by optimizing the ability of the hidden nodes to store the necessary weight information, while avoiding the over-determination that would result from an excessive number of units. While greater number of units would have given the model greater flexibility, it would have increased complexity at the cost of decreasing generalizability to the testing sample. Similarly, not enough units would not have produced a proper fit with the data and would have reduced the power of the model. Therefore, various models were developed in order to find the proper balance and maximize the predictive power for each model.

In all models, the training and testing samples were selected at random from the existing data and the proportions were adjusted in order to maximize the training sample while preserving the appearance of all detected patterns in the testing sample, so as to be able to appropriately test the model. Other parameters that were varied in order to maximize the performance of the networks were learning rate and momentum. The variations in the learning rate parameter allowed the control of the amount of weight and bias change during the training of the network. Different problem conditions find better solutions with different size of changes in the architecture of the network. Regarding the momentum, it was used to prevent the network from converging too early to a local minimum, and conversely to avoid overshooting the global minimum of the function; thus, it is important to avoid having a value which is too large for the momentum (it can overshoot), or too low (it can get stuck in a local minimum). Balancing these parameters maximizes the solution, and if correctly identified provide a stable and reliable solution as the ones that were found in this study.

4.4 Predictive contribution by categories of variables

Besides studying the contribution of each variable individually for each neural network developed to classify the various expected performance levels (low performers, high performers, and three performance groups simultaneously), the contribution of each category or set of variables (background, basic cognitive processes, total reaction times for WMC operations and attentional networks, and learning strategies/motivation) was analysed for each ANN developed, and the total predictive weight for each category of variables, as well as their average, was determined. Table 6 and Figure 4 show that in terms of predictive weight, the most important variables when estimating the levels of predicted GPA performance for all three groups simultaneously, are the background factors (e.g., socio-economic status proxy data, type of secondary school, occupation and education of parents, etc.), but when comparing the two extreme predicted performance groups, it is interesting to note that specific patterns involving different variables are evident for low and high expected academic performance: learning strategies/motivation had a stronger predictive weight for students expected to be in the lowest 33% of GPA performance; on the other hand, for students predicted to belong to the highest 33% of GPA performance, background variables and some of the cognitive processing variables were those carrying the most predictive weight.



Table 6

Comparative Predictive Weight Contribution for the Three Levels of Academic Performance by each of the Categories of Predictor Variables

	Low 33%	Mid 33	High 33%	Mean Predictive Weight of Each Area
Background	28.40%	25.40%	36.60%	30.13%
Basic Cognitive	21.50%	25.70%	23.20%	23.47%
Reaction Time total	18.90%	21.10%	16.60%	18.87%
Learning Strategies/Motivation	31.20%	27.80%	23.60%	27.53%
	100%	100%	100%	

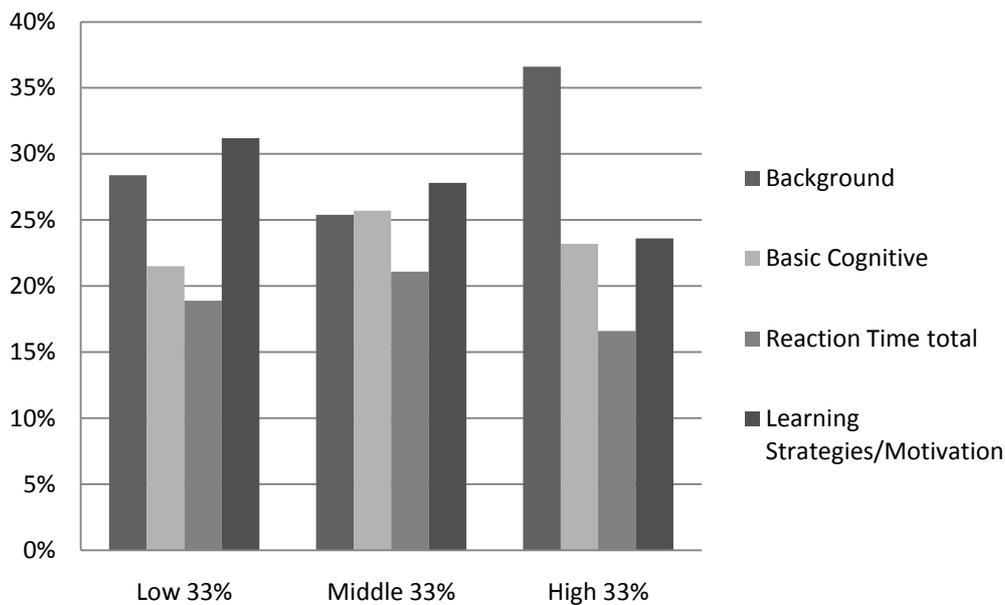
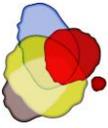


Figure 4. *Comparison of Predictive Weight Levels for the Three Levels of Academic Performance by Categories of Predictor Variables.*

4.5 Initial analysis of individual continuous estimates of future academic performance

While most of this study has been centered around the successful development of models to categorize expected levels of performance (which can be varied according to the problem situation), it is also important and useful to demonstrate that this machine learning approach can be used to predict individual specific outcomes (not just relatively broad performance categories). Although these performance categories can be very useful, as has been indicated for the identification and possible intervention in specific groups of high achievers or low achievers (i.e., learning disabilities, non-readiness for some specific task such as reading), and they can be used very effectively for targeted interventions in learning situations, it is also important to



be able to understand the underlying phenomenon at the individual level, considering performance a continuous variable.

For this reason, the predicted GPA-category (low-middle-high) probability values assigned by the network to each individual student were used to analyze their correlation with the observed GPA, as compared to the predicted value, in the context of the ANN₃ model, in which the whole sample of students was simultaneously classified in the three levels of expected performance. That is, the probability value for each student of belonging to a given category (all students received a certain probability of belonging to each of the outcome groups, as determined by the ANN), was correlated with the GPA actually obtained by each student. Results were indicative of a high degree of correlation between those measures.

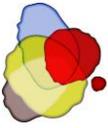
The three predicted groups of Low, Mid, and High performance had an actual observed GPA mean of 3.88 (SD = 1.21, n = 327), 5.67 (SD = .33, n = 243), and 7.28 (SD = .78, n = 294), respectively. All these average GPA means were significantly different from each other ($p < .000$). Within each one of the performance levels, the correlation of the ANN individual predicted value with the actual GPA was: Low 33%, $r = .78$; High 33%, $r = .73$, and for the whole sample of students, at all three levels, the correlation of the ANN predicted values with the observed GPA was $r = .86$. Further studies will continue to explore these individual relationships, but as they are, they confirm a high level of correlation between the actual GPA and the expected values assigned by the ANN.

4.6 Discriminant analyses (DA)

DA₁ focused on the attempted predictive classification of students expected to be in the lowest 33% of GPA average, compared to the rest of the students. One of the restrictions of this analysis has to do with the assumption of equality of covariance matrices that, in this case, is not violated (Box's $M = 5.253$, $F = .871$, $p = .515$). Gender, WMC and cognitive resources/learning strategies, were able to discriminate between the two groups of students, but not the rest of the variables, that were included in the ANN₁. The squared canonical correlation (CR^2) gives the amount of variation in the groups that is explained by the discriminating variables, which in this case was quite low (Wilk's $\lambda = .896$, $\chi^2 = 84.786$, $df = 3$, $p = .001$, $CR^2 = .323$).

DA₂ was carried out to attempt to discriminate between students expected to be in the highest 33% of GPA average, compared to the 67% of the rest of the students. The same independent variables that were used in the ANN₂ were entered in this analysis. Results show that the independent variables were not able to discriminate between both groups of students. The Box's M statistic is not significant (Box's $M = 11.813$, $F = .781$, $p = .700$), meaning that the assumption of equality of covariance matrices is not violated. In this analysis the squared canonical correlation indicated that the strength of the function is very low (Wilk's $\lambda = .926$, $\chi^2 = 58.694$, $df = 5$, $p = .001$, $CR^2 = .271$). Only gender, highest level of education of the father, WMC, and cognitive resources, and time management among the learning strategies set, were variables that entered significantly in this model.

DA₃ was carried out with the same variables as those used to develop ANN₃, in order to predict the expected GPA performance level of the three groups of academic performance simultaneously. The assumption of equality of covariance matrices was not violated (Box's $M = 7.522$, $F = .623$, $p = .824$). In this case, only gender, cognitive resources within the learning strategies set and WMC were significant for the model, and participated in the discrimination between the students in the three groups. But the model explained a very low and non-significant proportion of the variance (Wilk's $\lambda = .998$, $\chi^2 = 1.791$, $df = 2$, $p = .408$, $CR^2 = .048$).



5. Discussion and conclusions

The purpose of this study was to show the applicability and the effectiveness of the ANN approach to the predictive classification of students in the full range of academic performance (GPA), as well as to identify and understand the importance of the variables for each level (low, middle and high) of expected GPA. This methodology, using a predictive system, was chosen as it is very effective under conditions of very complex and great amount of data, in which a large number of variables interact in various complex and not very well understood patterns.

The results attained in this study have allowed the identification of the specific influence of each input set of variables on different levels of academic performance (high and low performance), on one hand, and common processes across all students, on the other hand. One important contribution of this predictive approach is the finding that the same variables have different effects in each group of students, defining specific patterns for each performance level. Although the contribution of each variable in a particular pattern carries a relatively small predictive weight, it is the combined effect of the pattern of variables which explains a lower or higher academic performance model.

Among the student group with the lowest 33% of academic performance, two main predictors are learning strategies components (cognitive resources/cognitive processing and time management). The importance of learning strategies as a mediating factor in a model predicting academic performance has been shown in different studies (Dupeyrat & Marine, 2005; Fenollar, et al., 2007; Simons et al., 2004; Weinstein & Mayer, 1986; Weinstein et al., 1987; Weinstein et al., 1982). However, this study added the contribution of a complex pattern of variables for a particular group of students, identifying specific learning strategies that help the classification of students in a low performance group (i.e., thoughts or behaviours that help to use imagery, verbal elaboration, organization strategies, and reasoning skills). Included in this set are learning strategies that help build bridges between what they already know, and what they are trying to learn and remember (i.e., knowledge acquisition, retention, and future application). In addition, variables related to speed of processing involved in WMC functioning have an important predictive weight for the determination and modelling of the low performance group. Other studies that have used ANN have also found that basic cognitive processing variables such as WMC and Executive Attention carried the most predictive weight in the low performance group of students (Kyndt et al., 2012, submitted; Musso & Cascallar, 2009a; Musso et al., 2012). Moreover, the literature has indicated the positive association between WMC and academic achievement (Gathercole, Pickering, Knight, & Stegmann, 2004; Riding, Grimley, Dahraei, & Banner, 2003). Regarding the relative importance of each variable, if we compare the relative role of WMC and other cognitive resources between the low and high performance groups, WMC and cognitive resources were far more important for lower GPA students. The fact that their importance for the prediction is much greater for the lower performing group is greatly due to the fact that all members of the high group had higher levels of WMC and cognitive resources, therefore not providing the necessary information to the network. On the other hand, it was an identifying characteristic of the low performing group which had consistently lower values of WMC and cognitive resources. Remediation programmes, tutorial systems and instruction methods should consider these specific learning strategies, cognitive processing characteristics and WMC resources, in order to provide basic support to students at risk. Such informed interventions would improve the possibilities of successful academic achievement for the at-risk groups, including those with particular learning difficulties.

Background variables together with reaction time measures and attentional executive control are the most important predictors for the highest academic performance group, as indicators of both efficiency in the processing and of adequate selection of information. Social background variables, such as educational level of the parents, have been found to be significant in a previous ANN study (Pinninghoff, Junemann et al., 2007), and these results have been replicated in this study. The executive control mechanism is responsible for resolving conflicts among responses (Fan et al., 2002). This attentional system has been closely related to working memory capacity (Redick & Engle, 2006), and was found to mediate and compensate WMC deficits for certain tasks (Musso et al., 2012). Other attentional networks seem to be much less discriminating among students who reach certain threshold levels needed for high academic performance. These findings have



significant implications in the way that the learning process can be addressed for students identified as potential high achievers. For this group, promoting learning through the use of metacognitive strategies, complex processing, and targeted teacher feedback would be an important way of maximizing their potential performance.

Regarding methodological implications, these results demonstrate the greater accuracy of the ANN approach compared to other traditional methods such as DA. Other studies have also made use of multilayer perceptron artificial neural networks, with positive results for the analysis of educational data (Abu Naser, 2012; Croy et al., 2008; Fong, et al., 2009; Kanakana, & Olanrewaju, 2011; Mukta & Usha, 2009; Ramaswami & Bhaskaran, 2010; Zambrano Matamala, et al., 2011). However, the present study has been able to maximize the precision obtained in the predictive classification of overall academic performance through the careful adjustment of network parameters and algorithms, producing highly accurate results with minimal misclassifications.

Similarly, the initial study of the correlation between the ANN probabilities of performance level assigned to each individual student, with the actual GPA observed, shows a significant degree of correlation between the two measures ($r = .86$ for the whole sample), with performance as a continuous variable. Further studies will refine the technique to maximize these individual results.

The results of the DA confirm the lack of significant linear relationships between the independent variables analysed in this study and academic performance. Neural network models have an important advantage in this respect, as they are able to model nonlinear and complex relationships among variables with greater precision and accuracy. Even though the assumptions required for traditional statistical predictive models (e.g. equality of covariance matrices) were not violated for the three stepwise discriminant analyses that were performed, the amount of variance explained was low in all three DA analyses. None of these analyses were able to discriminate with sufficient accuracy between the different levels of expected academic performance. When we compare these results with the ANNs modelled in this study, it can be concluded that ANNs are much more robust, and perform significantly better than other classical techniques, as prior studies have also indicated (Everson et al., 1994; Marquez et al., 1991).

This study has shown the power of this predictive approach using ANNs to model future overall academic performance in higher education, specifically in academic admissions and/or placement. To put the current results in perspective, if we consider one of the best known and most reliable tests currently in use, the SAT from The College Board, it has been found (Kobrin, Patterson, Shaw, Mattern, & Barbuti, 2008) that all sections of the SAT taken together, even with the more recent addition of a writing score, can predict at best 28% of the variance of the first-year college GPA for the average population of students. If we add to the SAT results the information of the GPA obtained in secondary education, the overall prediction is of only 38% of the variance of first-year college GPA (Kobrin et al, 2008). With the current ANN models, it has been possible to correctly classify 100% of student performance in the categories examined, that is, 100% of the students were correctly classified, and our research currently continues into the development of new predictive models, with much larger data sets, to classify students in much narrower bands of expected performance having already attained 98-99% accuracy in models for quintals of student performance distributions. In addition, work will also continue for the prediction of specific expected GPA results for each individual student.

In conclusion, the current predictive systems approach facilitates and maximizes the identification of those factors (or predictors) of the learning processes which participate in varying degrees in the modelling of different levels of performance in academic outcomes in higher education. If we can identify specific profiles of students, focusing on the most important variables, this opens major possibilities for the improvement of assessment procedures and the planning of pre-emptive interventions. Given that this methodology allows for the accurate prediction of actual academic performance at least one academic year in advance to it actually being measured (GPA), it has implications for the application of these methods in educational research and in the implementation of diagnostic “early-warning” programmes in educational settings. These results also inform cognitive theory and help in the development of improved automated tutoring and learning systems. Although some of the variables involved, such as educational level of the



parents, are impossible to alter in their effects on academic performance at the time of the assessment, they do inform policy and indicate the weight that many social and environmental factors influence future academic performance. This methodological and conceptual approach allows us to consider a large number of variables simultaneously and select those which are most relevant and allow a greater degree of intervention to improve student performance, including early intervention programmes for students in need of special support.

The capacity to very accurately classify expected student performance, which is also what tests attempt to do, without the performance sampling issues of traditional testing, and using a much broader spectrum of all factors influencing a student's overall performance, is a major advantage of the ANNs methodology. In fact, it also represents a more valid approach to educational assessment due to its overall accuracy and the breadth of the constructs considered to classify the expected performance. Traditional assessments are not sufficient for more complex assessments or for assessment systems that intend to serve multiple direct and indirect purposes, in complex educational situations (Mislevy, 2013; Mislevy, Steinberg, & Almond, 2003) In this respect, this new approach allows for the conceptualization and development of new modes of assessment which could facilitate breaking away from traditional forms of testing while at the same time improving the quality of the assessment process (Segers, Dochy & Cascallar, 2003).

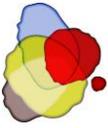
Finally, the use of ANN together with other methods as cluster analyses and Kohonen networks could contribute to the study of the specific patterns of those variables which influence the learning process for each level of performance. In fact, a major observation resulting from the data in this study is that variables contribute to the prediction in relatively small proportions, and it is the joint effect of many contributing variables that could cause significant changes in performance. In other words, there is no "magic bullet", rather the accumulation of effects from all these various sources that produces significant changes in outcomes. These results provide an insight into learning questions from a different perspective and one that has important implications for educational policy and education at large.

Keypoints

-  This approach provides a more contextualized and encompassing new mode of assessing expected performance without some of the pitfalls found in traditional forms of testing.
-  ANNs are a powerful tool to model future academic performance, specifically in academic diagnostic evaluations for placement and early-warning assessments.
-  This methodology demonstrates that variables impacting the outcome of the learning process are embedded in specific large-scale patterns which determine their degree of influence and direction of their effects.
-  A predictive systems approach is a valuable method to study the specific patterns of variables influencing the learning process at each level of expected performance, to better understand the determinants of learning outcomes and ways to improve them with early interventions.

References

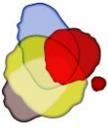
- Abu Naser, S. S. (2012). Predicting learners performance using artificial neural networks in linear programming intelligent tutoring system. *International Journal of Artificial Intelligence & Applications (IJAIA)*, 3(2), 65-73
- Anderson, J. R. (1983). *The architecture of cognition*. Cambridge, MA: Harvard University Press.
- Anderson, J. R. (1993). *Rules of the mind*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Anderson, J. R. (2002). Spanning seven orders of magnitude: A challenge for cognitive modeling. *Cognitive Science*, 26, 85–112.



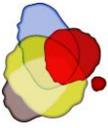
- Anderson, J. R. (2007) *How can the human mind occur in the physical universe?* New York: Oxford University Press.
- Anderson, J. R., Bothell, D., Byrne, M. D., Douglass S., Lebiere, C., & Yulin, Q. (2004). An integrated theory of the mind. *Psychological Review*, *111*(4), 1036–1060.
- Baddeley, A. D. (1986). *Working Memory*. Oxford: Clarendon Press.
- Bansal, A., Kauffman, R. J., & Weitz, R. R. (1993). Comparing the Modeling Performance of Regression and Neural Networks as Data Quality Varies: a Business Value Approach. *Journal of Management Information Systems*, *10*(1), 11- 32.
- Bekele, R., & McPherson, M. (2011). A Bayesian performance prediction model for mathematics education: A prototypical approach for effective group composition. *British Journal of Educational Technology*, *42*(3), 395–416.
- Biggs, J. (1987). *Study Process Questionnaire manual*. Melbourne, Australia: Australian Council for Educational Research.
- Birenbaum, M., Breuer, K., Cascallar, E., Dochy, F., Dori, Y., Ridgway, J., Wiesemes, R. (2006), & Nickmans, G. (Editor). A learning Integrated Assessment System. *Educational Research Review*, *1*, 61-67.
- Boekaerts, M., & Cascallar, E. (2006). How far have we moved toward the integration of theory and practice in Self-regulation? *Educational Psychology Review*, *18*(3), 199-210.
- Boekaerts, M. & Cascallar, E. C. (2011). Predicting and Explaining Writing Outcomes: Neural Network Methodology at work. Symposium: Predicting academic performance with the use of predictive systems analysis. *Proceedings of the Biennial Conference of the European Association for Research on Learning and Instruction (Earli)*. Exeter, UK, 30 August – 3 September 2011.
- Braten, I. & Stromso, H. (2006). Epistemological beliefs, interest, and gender as predictors of Internet-based learning activities. *Computers in Human Behavior*, *22*, 1027-1042.
- Cascallar, E. C., Boekaerts, M., & Costigan, T. E. (2006) Assessment in the Evaluation of Self- Regulation as a Process, *Educational Psychology Review*, *18*(3), 297-306.
- Cascallar, E. C., & Musso, M. F. (2008). Classificatory stream analysis in the prediction of expected reading readiness: Understanding student performance. *International Journal of Psychology, Proceedings of the XXIX International Congress of Psychology ICP 2008*, *43*(43/44), 231-.231.
- Castejón, J. L., & Navas, L. (1992). Determinantes del rendimiento académico en la educación secundaria. Un modelo causal. [Determinants of academic achievement in secondary education. A causal model]. *Análisis y Modificación de Conducta*, *18*(61), 697-728.
- Cattell, R. B. (1971). *Abilities: Structure, growth and action*. Boston: Houghton Mifflin.
- Chamorro-Premuzic, T., & Arteche, A. (2008). Intellectual competence and academic performance: preliminary validation of a model. *Intelligence*, *36*, 564-573.
- Colom, R., Escorial, S., Chun Shih, P., & Privado, J. (2007). Fluid intelligence, memory span, and temperament difficulties predict academic performance of young adolescents. *Personality and Individual Differences*, *42*, 1503-1514.
- Conway, A. R. A., Cowan, N., Bunting, M. F., Theriault, D., & Minkoff, S. (2002). A latent variable analysis of working memory capacity, short term memory capacity, processing speed, and general fluid intelligence. *Intelligence*, *30*, 163- 183.
- Conway, A. R. A., & Engle, R.W. (1996). Individual differences in working memory capacity: More evidence for a general capacity theory. *Memory*, *4*, 577-590.
- Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review*, *12*(5), 769-786
- Croy, M., Barnes, T., & Stamper, J. (2008). Towards an intelligent tutoring system for propositional proof construction. In A. Briggle, K. Waelbers, and P. Brey (Eds.), *Computing and Philosophy* (pp. 145-215). Amsterdam, The Netherlands: IOS Press.
- Daneman, M., & Carpenter, P. A. (1980). Individual-differences in working memory and reading. *Journal of Verbal Learning and Verbal Behaviour*, *19*, 450 - 466.



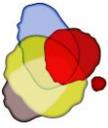
- Detienne, K. B., Detienne, D. H., & Joshi, S. A. (2003). Neural networks as statistical tools for business researchers. *Organizational Research Methods*, 6, 236-265.
- Duliba, K. A. (1991) Contrasting Neural Nets with Regression in Predicting Performance in the Transportation Industry. *Proceedings of the Twenty-Fourth Annual Hawaii International Conference on System Sciences*, 4.
- Dupeyrat, C., & Marine, C. (2005). Implicit theories of intelligence, goal orientation, cognitive engagement, and achievement: A test of Dweck's model with returning to school adults. *Contemporary Educational Psychology*, 30(1), 43-59.
- Engle, R.W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science*, 11, 19-23.
- Engle, R.W., & Kane, M. J. (2004). Executive attention, working memory capacity, and a two-factor theory of cognitive control. In B. Ross (Ed.), *The Psychology of Learning and Motivation* (pp. 145-199). New York, NY: Elsevier.
- Eriksen, B. A., & Eriksen, C.W. (1974). Effects of noise letters upon the identification of a target letter in a non search task. *Perception and Psychophysics*, 16, 143-149.
- Everson, H. T. (1995). Modelling the student in intelligent tutoring systems: The promise of a new psychometrics. *Instructional Science*, 23(5-6), 433-452.
- Everson, H. T., Chance, D., & Lykins, S. (1994). Exploring the use of artificial neural networks in educational research. *Paper presented at the annual meeting of the American Educational Research Association*, New York.
- Fan, J., McCandliss, B. D., Summer, T., Raz, A., & Posner, M.I. (2002). Testing the efficiency and independence of attentional networks. *Journal of Cognitive Neuroscience*, 14(3), 340-347.
- Feldman Barrett, L., Tugade, M. M., & Engle, R. W. (2004). Individual differences in working memory capacity and dual-process theories of mind. *Psychological Bulletin*, 130, 553-573.
- Fenollar, P., Roman, S., & Cuestas, P. J. (2007). University students' academic performance: An integrative conceptual framework and empirical analysis. *British Journal of Educational Psychology*, 77, 873-891.
- Fernandez-Castillo, A., & Gutiérrez-Rojas, M. E. (2009). Selective attention, anxiety, depressive symptomatology and academic performance in adolescents. *Electronic Journal of Research in Educational Psychology*, 7(1), 49-76.
- Fletcher, J. M. (2005). Predicting math outcomes: Reading predictors and comorbidity. *Journal of Learning Disabilities*, 38(4), 308-312.
- Fong, S., Si, Y.-W., & Biuk-Aghai, R. P. (2009). Applying a Hybrid Model of Neural Network and Decision Tree Classifier for Predicting University Admission. *Proceedings of the 7th International Conference on Information, Communication, and Signal Processing (ICICS2009)*, pp. 1-5, Macau, China, IEEE Press.
- Garson, G. D. (1998). *Neural Networks. An Introductory Guide for Social Scientists*. London: Sage Publications Ltd.
- Gathercole, S. E., Pickering, S. J., Knight, C., & Stegmann, Z. (2004). Working memory skills and educational attainment: Evidence from national curriculum assessments at 7 and 14 years of age. *Applied Cognitive Psychology*, 18, 1-16.
- Gazzaniga, M., Ivry, R., & Mangun, G. (2002). *Cognitive neuroscience: The biology of the mind (2nd ed.)*. New York, NY: W.W. Norton
- Grimley, M., & Banner, G. (2008). Working memory, cognitive style, and behavioural predictors of GCSE exam success. *Educational Psychology*, 28(3), 341-351.
- Grossberg, S. (1980). How does the brain build a cognitive code? *Psychological Review*, 87, 1- 51.
- Grossberg, S. (1982). *Studies of mind and brain: Neural principles of learning, perception, development, cognition and motor control*. Boston: Reidel Press.
- Gsanger, K., W., Homack, S., Siekierski, B., & Riccio, C. (2002). The relation of memory and attention to academic achievement in children. *Archives of Clinical Neuropsychology*, 17(8), 790.
- Hailikari, T., Nevgi, & A., Komulainen, E. (2008). Academic self-beliefs and prior knowledge as predictors of student achievement in Mathematics: a structural model. *Educational Psychology*, 28(1), 59-71.



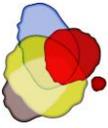
- Hardgrave, B. C., Wilson, R. L., & Walstrom, K. A. (1994). Predicting Graduate Student Success: A Comparison of Neural Networks and Traditional Techniques. *Computer and Operations Research*, 21(3), 249-263.
- Hazy, T. E., Frank, M. J., & O' Reilly, R. C. (2006). Banishing the Homunculus: Making Working Memory Work. *Neuroscience* 139, 105–118.
- Heitz, R. P., Redick, T. S., Hambrick, D. Z., Kane, M. J., Conway, A. R. A., & Engle, R. W. (2006). Working memory, executive function, and general fluid intelligence are not the same. *Behavioral and Brain Sciences*, 29, 135-136.
- Jarrold, C., & Towse, J. N. (2006). Individual differences in working memory. *Neuroscience*, 139, 39-50.
- Jimmerson, S. R., Dubrow, E. H., Adam, E., Gunnar, M., & Bozoky, I. K. (2006). Associations among academic achievement, attention, and andrenocortical reactivity in Caribbean village children. *Canadian Journal of School Psychology*, 21, 120-138.
- Kanakana, G., & Olanrewaju, A. (2011). Predicting student performance in engineering education using an artificial neural network at Tshwane University of Technology, *Proceedings of the ISEM*, Stellenbosch, South Africa.
- Kane, M. J., Hambrick, D. Z., Tuholski, S.W., Wilhelm, O., Payne, T.W., & Engle, R.W. (2004). The generality of working memory capacity: A latent variable approach to verbal and visuospatial memory span and reasoning. *Journal of Experimental Psychology: General*, 133, 189-217.
- Kent, R. (2009). Rethinking data analysis – part two. Some alternatives to frequentist approaches. *International Journal of Market Research*, 51, 181-202.
- Kobrin, J. L., Patterson, B. F., Shaw, E. J., Mattern, K. D., & Barbuti, S. M. (2008). Validity of the SAT for predicting first-year college grade point average. *College Board Research Report 2008-5*. New York: The College Board. Retrieved from <http://research.collegeboard.org/rr2008-5.pdf>.
- Kohavi, R. & Provost, F. (1998). Glossary of terms. *Machine Learning*, 30(2–3): 271–274.
- Kuncel, N. R., Hezlett, S. A., & Ones, D. S. (2001). A comprehensive meta-analysis of the predictive validity of the Graduate Record Examinations: Implications for graduate student selection and performance. *Psychological Bulletin*, 127(1), 162-181.
- Kuncel, N. R., Crede, M., Thomas, L. L., Klieger, D.M., Seiler, S.N., & Woo, S.E. (2004). A meta-analysis of the Pharmacy College Admission Test (PCAT) and grade predictors of pharmacy student success. *Annual conference of the American Psychological Society*, Chicago, IL.
- Kuncel, N. R., Hezlett, S. A., & Ones, D. S. (2004). Academic performance, career potential, creativity, and job performance: Can one construct predict them all? *Journal of Personality and Social Psychology*, 86(1), 148-161.
- Kuncel, N. R., Crede, M., Thomas, L. L., Klieger, D. M., Seiler, S. N., & Woo, S. E. (2005). A meta-analysis of the Pharmacy College Admission Test (PCAT) and grade predictors of pharmacy student success. *American Journal of Pharmaceutical Education*, 69(3), 339-347.
- Krumm, S., Ziegler, M., Buehner, M. (2008). Reasoning and working memory as predictors of school grades. *Learning and Individual Differences*, 18 (2), 248-257.
- Kyllonen, P. C., & Christal, R. E. (1990). Reasoning ability is (little more than) working-memory capacity?! *Intelligence*, 14, 389-433.
- Kyllonen, P. C., & Stephens, D. L. (1990). Cognitive abilities as determinants of success in acquiring logic skill. *Learning and Individual Differences*, 2, 129-160.
- Kyndt, E., Cascallar, E., & Dochy, F. (2012). Individual differences in working memory capacity and attention, and their relationship with students' approaches to learning. *Higher Education*, 64(3), 285-297.
- Kyndt, E., Musso, M., Cascallar, E., & Dochy, F. (2012, Submitted). Predicting academic performance: The role of cognition, motivation and learning approaches. A neural network analysis. *Journal of Further and Higher Education*.
- Landerl, K. (2010). Temporal processing, attention, and learning disorders. *Learning & Individual Differences*, 20(5), 393-401.
- Linn, R. L., & Hastings, C. N. (1984). A meta-analysis of the validity of predictors of performance in law school. *Journal of Educational Measurement*, 21, 245-259.



- Lippman, R. (1987). An introduction to computing with neural nets. *IEEE ASSP Magazine*, 3(4), 4-22.
- Lovett, M. W. (1979). The selective encoding of sentential information in normal reading development. *Child Development*, 50(3), 897.
- Lykins, S., & Chance, D. (1992). Comparing artificial neural networks and multiple regression for predictive application, *Proceedings of the Eight Annual Conference on Applied Mathematics*, Edmond OK, 155-169
- Marquez, L., Hill, T., Worthley, R., & Remus, W. (1991). Neural network models as an alternative to regression. *Proceedings of the IEEE 24th Annual Hawaii International Conference on Systems Sciences*, 4, 129-135.
- Marshall, D. B., & English, D. J. (2000). Neural network modelling of risk assessment in child protective services. *Psychological Methods*, 5(1), 102-124.
- Maucieri, L. P. (2003). Predicting behavior with an artificial neural network: A comparison with linear models of prediction (January 1, 2003). *ETD Collection for Fordham University*, NY, USA. Retrieved from <http://fordham.bepress.com/dissertations/AAI3098134>.
- Mavrovouniotis, M. L. & Chang, S. (1992). Hierarchical neural networks. *Computers & Chemical Engineering*, 16(4), 347-369.
- Miñano, P., Gilar, R., & Castejón, J. L. (2012) A structural model of cognitive-motivational variables as explanatory factors of academic achievement in Spanish Language and Mathematics. *Anales de Psicología*, 28(1), 45-54.
- Mislevy, R. J. (2013). Measurement is a Necessary but not Sufficient Frame for Assessment. *Measurement*, 11, 47-49, 2013
- Mislevy, R. J., Steinberg, L. S., & Almond, R. A. (2003). On the structure of educational assessments. *Measurement: Interdisciplinary Research and Perspectives*, 1, 3-67.
- Mukta, P., & Usha, A., (2009). A study of academic performance of business school graduates using neural network and statistical techniques. *Expert Systems with Applications*, 36(4), 7865-7872.
- Musso, M. F., & Cascallar, E. C. (2009a). New approaches for improved quality in educational assessments: Using automated predictive systems in reading and mathematics. *Journal of Problems of Education in the 21st Century*, 17, 134-151.
- Musso, M. F., & Cascallar, E. C. (2009b). 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 (Volume I)*, (pp. 433-459). Argentina: CIIPME/CONICET.
- Musso, M. F., Kyndt, E., Cascallar, E. C., & Dochy, F. (2012). Predicting mathematical performance: The effect of cognitive processes and self-regulation factors. *Education Research International*. Vol. 12.
- Nasr, G. E., Badr, E. A., & Joun, C. (2002). Cross Entropy Error Function In Neural Networks: Forecasting Gasoline Demand. *FLAIRS-02 Proceedings of the AAAI*. Retrieved from <http://www.aaai.org/Papers/FLAIRS/2002/FLAIRS02-075.pdf>
- Navas, L., Sampascual, G., & Santed, M. A. (2003). Predicción de las calificaciones de los estudiantes: la capacidad explicativa de la inteligencia general y de la motivación. [Prediction of students' performance scores: the role of the general intelligence and motivation. *Journal of General and Applied Psychology*], 56(2), 225-237.
- Neal, W., & Wurst, J. (2001). Advances in market segmentation. *Marketing Research*, 13(1), 14-18.
- Passolunghi, M. C., & Pazzaglia, F. (2004). Individual differences in memory updating in relation to arithmetic problem solving. *Learning and Individual Differences* 14(4), 219-230.
- Perkins, K., Gupta, L. & Tammana (1995). Predict item difficulty in a reading comprehension test with an artificial neural network. *Language Testing*, 12(1), 34-53.
- Pickering, S. J. (2006). *Working memory and education*. USA: Academic Press.
- Pinninghoff Junemann, M. A., Salcedo Lagos, P. A., & Contreras Arriagada, R. (2007). Neural networks to predict schooling failure/success. In J. Mira & J.R. Alvarez (Eds.), *IWINAC 2007, Part II, LNCS 4528*(pp. 571-579). Berlin / Heidelberg: Springer-Verlag.
- Pintrich, P. R. (2000). The role of goal orientation in self-regulated learning. In M. Boekaerts, P.R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 452-502). San Diego, CA: Academic Press.



- Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology*, 41A, 19-45.
- Posner, M. I., & Petersen, S. E. (1990). The attention system of the human brain. *Annual Review Neuroscience*, 13, 25-42.
- Posner, M. I., & Rothbart, M. K. (1998). Attention, self-regulation and consciousness. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 353, 1915–1927.
- Ramaswami, M. M., & Bhaskaran, R. R. (2010). A CHAID based performance prediction model in educational data mining. *International Journal of Computer Science Issues*, 7(1), 10-18.
- Redick, T. S., & Engle, R.W. (2006). Working memory capacity and attention network test performance. *Applied Cognitive Psychology*, 20, 713-721.
- Reid, R. (2006). Self-regulated strategy development for written expression with students with attention deficit/ hyperactivity disorder. *Exceptional Children*, 73(1), 53-67.
- Riccio, C. A., Lee, D., Romine, C. Cash, D., & Davis, B. (2002). Relation of memory and attention to academic achievement in adults. *Archives of Clinical Neuropsychology*, 18(7), 755-756.
- Riding, R. J., Grimley, M., Dahraei, H., & Banner, G. (2003). Cognitive style, working memory and learning behaviour and attainment in school subjects. *British Journal of Educational Psychology*, 73, 749-769.
- Roth, P. L., Be Vier, C. A., Switzer, F. S., & Schippmann, J. S. (1996). Meta-analyzing the relationship between grades and job performance. *Journal of Applied Psychology*, 81, 548-556.
- Roth, P. L., & Clarke, R. L. (1998). Meta-analyzing the relation between grades and salary. *Journal of Vocational Behavior*, 53, 386-400.
- Ruban, L. M., & McCoach, D. B. (2005). Gender differences in explaining grades using structural equation modeling. *The Review of Higher Education*, 28, 475-502.
- Rueda, M. R., Posner, M. I., & Rothbart, M. K. (2004). Attentional control and self regulation. In R.F. Baumeister & K.D. Vohs (Eds), *Handbook of Self Regulation: Research, Theory, and Applications*, New York: Guilford Press, 14: 283-300.
- Rumelhart, D., Hinton, G. & Williams, R. (1986). Learning representations by back-propagating errors. *Nature*, 323, 533- 536.
- Rumelhart, D. E., McClelland, J. L., & the PDP research group. (1986). *Parallel distributed processing: Explorations in the microstructure of cognition. Volume I*. Cambridge, MA: MIT Press.
- Schmidt, F. L. (2002). The role of general cognitive ability and job performance: Why there cannot be a debate. *Human Performance*, 15, 187–210.
- Segers, M., Dochy, F., & Cascallar, E. (2003). *Optimizing new modes of assessment: In search of qualities and standards*. The Netherlands: Kluwer Academic Publishers.
- Simons, J., Dewitte, S., & Lens, W. (2004). The role of different types of instrumentality in motivation, study strategies, and performance: Know why you learn, so you'll know what you learn! *British Journal of Educational Psychology*, 74, 343-360.
- Snyderman, M., & Rothman, S. (1987). Survey of expert opinion on intelligence and aptitude testing. *American Psychologist*, 42(2), 137-144
- Specht, D. (1991). A general regression neural network. *IEEE transactions on neural networks*, 2(6), 568-576.
- St Clair-Thompson, H. L., & Gathercole, S. E. (2006). Executive functions and achievements in school: Shifting, updating, inhibition, and working memory. *The Quarterly Journal of Experimental Psychology*, 59(4), 745-759.
- Strucchi, E. (1991). *Inventario de Estrategias de Aprendizaje y de Estudio. [Learning Strategies Inventory and Study]*. Buenos Aires: Psicoteca.
- Turner, E. A., Chandler, M., & Heffer, R. W. (2009). Influence of parenting styles, achievement motivation, and self-efficacy on academic performance in college students. *Journal of College Student Development*, 50, 3, 337-346.
- Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods*, 37(3), 498-505.
- Vandamme, J. P., Meskens, N., & Superby, J. F. (2007). Predicting academic performance by data mining methods. *Education Economic*, 15(4), 405-41.



- Walczak, S. (1994). Categorizing university student applicants with neural networks. *IEEE International conference on neural networks*, 6, 3680-3685.
- Weinstein, C. E., & Mayer, R.E. (1986). The teaching of learning strategies. In M.C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed.). Macmillan, New York.
- Weinstein, C. E. & Palmer, D. R. (2002). *LASSI: User's Manual (2nd Edition)*. Clearwater, FL: H&H Publishing Company, Inc.
- Weinstein, C. E., Palmer, D. R., & Schulte, A. C. (1987). *Learning and study strategies inventory*. Clearwater, FL: H & H Publishing company, Inc.
- Weinstein, C. E., Schulte, A. C., & Cascallar, E. C. (1982). The learning and studies strategies inventory (LASSI): Initial design and development. *Technical Report, US Army Research Institute for the Social and Behavioural Sciences*, Alexandria, VA.
- Weiss, S. M. & Kulikowski, C. A. (1991). *Computer systems that learn*. San Mateo, CA: Morgan Kaufmann Publishers.
- Welsh, M.C., Satterlee-Cartmell, T., & Stine, M. (1999). Towers of Hanoi and London: Contribution of working memory and inhibition to performance. *Brain Cognition*, 41(2), 231-242.
- White, H. & Racine, J. (2001): Statistical inference, the bootstrap, and neural network modelling with application to foreign exchange rates. *IEEE Transactions on Neural Networks: Special Issue on Neural Networks in Financial Engineering*, 12, 657-673.
- Wilson, R. L. & Hardgrave, B. C. (1995). Predicting graduate student success in a MBA program: Regression vs. classification. *Educational and Psychological Measurement*, 55, 186-195.
- Zambrano Matamala, C., Rojas Díaz, D., Carvajal Cuello, K., & Acuña Leiva, G. (2011). Análisis de rendimiento académico estudiantil usando data warehouse y redes neuronales. [Analysis of students' academic performance using data warehouse and neural networks] *Ingeniare. Revista Chilena de Ingeniería*, 19(3), 369-381.
- Zeegers, P. (2004). Student learning in higher education: A path analysis of academic achievement in science. *Higher Education Research & Development*, 23(1), 35-56.