MIND, BRAIN, AND EDUCATION

Modeling the Contribution of Genetic Variation to Cognitive Gains Following Training with a Machine Learning Approach

Mariel F. Musso[1,2,3], Lina M. Cómbita[1], Eduardo C. Cascallar[4], and M. Rosario Rueda[1]

ABSTRACT— The objective of this research was to develop robust predictive models of the gains in working memory (WM) and fluid intelligence (Gf) following executive attention training in children, using genetic markers, gender, and age variables. We explore the influence of genetic variables on individual differences in susceptibility to intervention. Sixty-six children (males: 54.2%) aged 50.9–75.9 months participated in a four-weeks computerized training program. Information on genes involved in the regulation of dopamine, serotonin, norepinephrine, and acetylcholine was collected. The standardized pre- to post-training gains of two dependent measures were considered: WM Span backwards condition (WISC-III) and the IQ-f factor from the Kaufman Brief Intelligence Test (K-BIT). A machine-learning methodology was implemented utilizing multilayer perceptron artificial neural networks (ANN) with a backpropagation algorithm. Both ANN models reached high overall accuracy in their predictive classification. Variations in genes involved in dopamine and norepinephrine neurotransmission affect children's susceptibility to benefit from executive attention training, a pattern that is consistent with previous studies.

[1]Department of Experimental Psychology and Center for Research on Mind, Brain and Behavior (CIMCYC), Universidad de Granada
[2]Interdisciplinary Center for Research in Mathematical and Experimental Psychology (CIIPME), National Council for Scientific and Technical Research (CONICET)
[3]Faculty of Health Sciences, Department of Psychology, UADE
[4]Faculty of Psychology and Educational Sciences, KU Leuven

Address correspondence to Mariel F. Musso, Department of Experimental Psychology and Center for Research on Mind, Brain and Behavior (CIMCYC), Universidad de Granada, C/Campus de la Cartuja S/N, 18011, Granada; e-mail: mariel.musso@ugr.es

Mariel F. Musso and Lina M. Cómbita shared first-authorship.
Eduardo C. Cascallar and M. Rosario Rueda shared supervision.

The field of cognitive training has had significant activity in the last decade. Work in this area has involved a broad range of executive functions, such as executive attention, working memory (WM), reasoning, and shifting of attention during development (Kerns, Eso, & Thomson, 1999; Rueda, Rothbart, McCandliss, Saccomanno, & Posner, 2005; Thorell, Lindqvist, Nutley, Bohlin, & Klingberg, 2009).

According to Posner's neurocognitive model, executive attention is a functional network essential for the regulation of thoughts, emotions, and action, and the effective management of conflict between different response options or sources of stimulation (Posner & Petersen, 1990). The executive attention network involves the activation of two different sets of structures in the brain: the so-called frontoparietal control system and the cingulo-opercular system (Petersen & Posner, 2012). Empirical evidence has shown that individual differences in executive attention are related to very important outcomes during childhood such as school competence and socialization (Rueda, Checa, & Rothbart, 2010). Executive attention is thought to be a core process enabling superior cognitive skills such as fluid reasoning and the voluntary regulation of actions (Rueda, 2018). Thus, training of executive attention often results in transfer to fluid reasoning (Rueda et al., 2005; Pozuelos et al., 2019) as well as improved performance of a range of executive processes involving self-regulation (Rueda, Checa, & Cómbita, 2012). Further, there seems to exist quite a degree of transfer of training among different executive processes, such as attention, WM, inhibitory control or cognitive flexibility, particularly in children populations (Karbach & Kray, 2009; Thorell et al., 2009).

WM refers to a limited capacity system responsible for the active maintenance and manipulation of information available to the cognitive system (Conway, Kane, & Al, 2005). WM is necessary for the performance of complex cognitive processes such as learning, comprehension, and

reasoning (Arteaga Díaz & Pimienta Jiménez, 2006), and it is related to important cognitive abilities and outcomes such as fluid intelligence (Gf) (Engle Laughlin, Tuholski, & Conway, 1999; Kane et al., 2004; Shipstead, Harrison, & Engle, 2016), math performance (Engle & Kane, 2003; Musso, Kyndt, Cascallar, & Dochy, 2012; Musso, 2016), and multitasking (Hambrick, Oswald, Darowski, Rench, & Brou, 2010).

A broad set of evidence (Heitz, Unsworth, & Engle, 2005; Kane et al., 2004; Kane, Conway, Hambrick, & Engle, 2007) has shown that both executive attention and WM are highly related to Gf, the capacity to find relations/patterns, and to infer rules for novel problems (Horn & Cattell, 1966). Moreover, the activation of brain areas associated with the executive attention network highly overlap with those that support general intelligence (Duncan & Owen, 2000). Even though Gf works independently from acquired knowledge, it has a key role for academic achievement (Alloway & Alloway, 2010; Lynn, Meisenberg, Mikk, & Williams, 2007).

Given the importance of executive cognitive processes for the development of children, an important field of research has aimed at understanding the impact of cognitive training on a variety of children's outcomes. Regarding executive attention training, several studies have found that Gf improves after training executive attention either directly on the targeted population (Karbach & Kray, 2009; Minear & Shah, 2008; Rueda et al., 2012) or using a combination of training children's attention with a family-based intervention aimed at improving parental regulatory skills for themselves and the children (Neville et al., 2013). This generalization to nontrained abilities has been found to be enhanced if the training is provided along with a metacognitive scaffolding designed to increase children's metacognitive knowledge and strategies for the tasks (Pozuelos et al., 2019). This far transfer effect to Gf measures appears to be maintained beyond the training period (2 months later) without further training in preschool children (Rueda et al., 2012). In addition, executive attention training has also shown to positively impact the functioning of the neural areas and networks that underlie the trained cognitive processes, improving both speed and efficiency of the executive attention network (Rueda et al., 2005; Rueda et al., 2012).

As executive attention, WM and Gf also encompass a set of interrelated functions and overlapping neural substrates. The rationale behind cognitive training targeting different executive functions is that improving executive attention and/or WM should lead to better performance not only in similar tasks (near transfer) but also to tasks that tap cognitive skills that have not been directly trained (far transfer) such as reasoning. In this regard, there is some evidence showing that WM training could lead to significant improvements of children's WM performance after training (Jaeggi, Buschkuehl, Jonides, & Shah, 2011; Klingberg, 2010). However, the extent to which such effects are transferred to other domains such as reasoning or school performance is still controversial (Melby-Lervåg & Hulme, 2012; Redick, 2015; Redick, Melby-Lervåg, & Hulme, 2016; Shipstead, Redick, & Engle, 2012; Unsworth, Redick, Heitz, Broadway, & Engle, 2009).

Researchers in the field of cognitive training have turned their attention toward individual differences that could explain these inconsistencies. One of the main questions is whether participant's skills prior to the implementation of a training program influence the extent to which they can benefit from the training. However, evidence in this regard is still inconclusive. While some authors have found that people with deficits in WM and/or Gf can benefit to a greater extent from a cognitive training program (Diamond & Lee, 2011; Jaeggi, Buschkuehl, Jonides, & Perrig, 2008), others have reported that individuals with high WM capacity show larger training-related gains than those with a low WM capacity (Fossella, Posner, Fan, Swanson, & Pfaff, 2002; Foster, 2017; Fuchs et al., 2013; Swanson, Moran, Lussier, & Fung, 2014).

Another focus of study addressing individual differences have aimed at understanding how certain individual variables can modulate the extent to which children benefit from training. While some of those variables are related to children's performance during training (Jaeggi et al., 2011; Söderqvist, Bergman Nutley, Ottersen, Grill, & Klingberg, 2012), other variables are related to children's constitutional factors such as temperament (Studer-Luethi, Bauer, & Perrig, 2016) or genetic endowment (Bäckman & Nyberg, 2013).

Among genetic-based individual differences, variations in specific dopamine-related genes that are involved in regulating the availability of this neurotransmitter in the prefrontal cortex and the striatum have been linked to variations in cognitive training effects. Specifically, polymorphisms within the gene coding for the dopamine transporter protein (DAT1) have been associated with improved effects of training on Gf and WM (Söderqvist et al., 2012). Likewise, (Brehmer, Westerberg, Bellander, Fürth, & Karlsson, 2009) found that carriers of at least one copy of the 9-repeat allele of the DAT1 gene show larger training effects in visuospatial WM compared to participants homozygous for the 10-repeat allele. The dopamine transporter is the main mechanism of dopamine regulation in the synaptic cleft within the striatum. Therefore, it has been argued that the different concentrations of DAT proteins associated to different alleles within the gene, might be the molecular basis of neural plasticity through its influence on the activity of dopamine D1 and D2 receptors (Söderqvist et al., 2012).

Other studies have found that polymorphisms associated to different levels of the catechol O-methyltransferase (COMT) enzymatic activity can also influence training-related gains in different cognitive domains. However,

variations associated to better cognitive gains appear to differ depending on the age of the population included in the study. For instance, two independent studies have found an association between the Val allele of the gene and increased WM plasticity in adults (Bellander et al., 2015; Colzato, van den Wildenberg, & Hommel, 2014), while others have reported better cognitive gains for carriers of the Met allele, particularly in younger participants (Zhao et al., 2020). These results corroborate the central role of dopamine availability in the prefrontal cortex for the plasticity of cognitive processes associated with cognitive flexibility, such as those involved in the updating component of WM or the top-down orientation and reorientation of attention associated to executive attention.

The interaction of genes and experience has also been widely reported evidencing how specific genetic variations moderate the impact of environmental variables on children's cognitive performance and behavior. For example, the presence of the 7-repeat allele, a variation of the DRD4 dopamine receptor gene, has been shown to interact with the quality of parenting in early childhood to influence children's temperamental characteristics related to sensation seeking, including impulsivity, high intensity pleasure and activity level (Sheese, Voelker, Rothbart, & Posner, 2007). Similarly, (Bakermans-Kranenburg & Van Ijzendoorn, 2006) found that children' susceptibility to maternal insensitivity and its impact on externalizing behaviors was highly dependent on the presence of the DRD4 7-repeat variation. In addition, (Voelker, Sheese, Rothbart, & Posner, 2009) have observed that both the COMT genotype and haplotype (a combination of alleles within the COMT gene locus) interact with parenting quality, influencing the performance of 2-year-old children in a visual sequence task that targets attention.

Similar results have been found with genes related to the modulation of different neurotransmitters in areas of the brain associated with executive attention and other executive processes (Posner, Rothbart, & Sheese, 2007). For instance, (Kochanska, Philibert, & Barry, 2009) reported that the development of self-regulation during the preschool years could be partially explained by the interaction between a polymorphism in the serotonin transporter (5HTTPR) and the quality of early mother–child attachment. Additional evidence from Caspi et al. (2002), shows that variations within the monoamine oxidase A (MAOA gene) that codes for the isoenzyme MAOA which regulates the levels of serotonin, norepinephrine, and dopamine, modulates the influence of child maltreatment on the development of antisocial behavior and conduct disorders (Caspi et al., 2002).

Clearly, significant evidence has been provided in this field of study, showing that genetic variations can moderate the extent to which children can benefit from cognitive training. However, as new discoveries are made in the field, new and relevant questions start to emerge. For instance, it is important to discern whether it is possible to predict which children could benefit from a cognitive intervention considering only specific genetic patterns as predictors. Given the relevance of this question, in the present study, we aimed at developing predictive models of children's cognitive-related gains in measures of WM and Gf following executive attention training. In addition, this study aims at understanding the contribution of specific genetic markers to these predictive models, and to compare particular patterns predicting gains in WM and Gf.

Some authors have shown that traditional statistical methods do not always yield accurate predictions and/or classifications (Bansal, Kauffman, & Weitz, 1993; Duliba, 1991; Everson, 1995). A more robust and accurate approach has been developed during the last decade and applied in health and education fields for the purpose of prediction (Cascallar, Boekaerts, & Costigan, 2006; Everson, Chance, & Lykins, 1994; Gorr, 1994; Hardgrave, Wilson, & Walstrom, 1994; Musso & Cascallar, 2009; Musso et al., 2012; Musso, Kyndt, Cascallar, & Dochy, 2013; Musso, Cascallar, Bostani, & Crawford, 2020; Musso, Hernández, & Cascallar, 2020). Machine-learning techniques, such as methods using artificial neural networks (ANN), have been shown to be very effective to study problems consisting of a large number of variables in complex, nonlinear, and poorly understood interactions (Cascallar, Musso, Kyndt, & Dochy, 2014). In addition to being powerful classifiers, ANN build plausible architectures to explore the participation of variables involved in the modeling of a problem (Detienne, Detienne, & Joshi, 2003; Neal & Wurst, 2001; White & Racine, 2001). This methodological approach could allow us to consider simultaneously a large number of factors without the usual parametric constraints, and it could help us understand whether genetic-based individual differences can explain differences in children's susceptibility to training. These advantages allow for a better understanding of the genetic factors that could be related to outcomes associated with cognitive training, taking into account their complex interactions. Moreover, detecting children who can benefit from a systematic cognitive training intervention might inform the development of more targeted program designs.

PRESENT STUDY

The present study is based on the effects of cognitive training demonstrated in a previous study (Pozuelos et al., 2019). The main research question is whether it is possible to predict cognitive gains in children, with high accuracy, considering only specific genetic variants. The answer to this question will lead us to the second research question of this study: which are the most important patterns of predictors contributing to the classification between children who benefit or do not benefit from training?

Given the evidence in the field, we hypothesized that ANN can, in fact, help us to predict cognitive training gains in preschool children, an age group that shows high levels of individual differences. We also hypothesized that those children carrying genetic variations associated with a more efficient neural functioning within the executive attention network will show larger training benefits. Although machine learning techniques have been widely used in several different fields of research and in applied environments, to our knowledge, there are no known studies applied with this objective using genetic data during early child development.

METHOD

Participants

The data base consists of a sample of 66 children corresponding to the training group of a previous study (Pozuelos et al., 2019), male = 54.2%, ages between 50.9 and 75.9 months ($M = 63.07$; $SD = 7.31$), from Granada (Spain). Socioeconomic status, based on the educational level of the mother, was not statistically significantly different between the various genotype groups. Inclusion criteria were that children had normal or corrected-to-normal sensory capacities, no history of chronic illness and/or psychopathologies, and were not under pharmacological treatment of any kind as informed by caregivers. Ethical approval was obtained from the Research Ethics Committee of the University of Granada. Parents were informed about the purpose of the study, and they gave written consent to participate in the study (for a more detailed description see Pozuelos et al., 2019).

MATERIALS AND PROCEDURE

Training Program

A computerized training program was administered for 4 weeks (ten sessions: 45 min per session). The program consisted of 14 computerized exercises divided into five categories: (1) Tracking/Anticipatory; (2) Attention Focusing/Discrimination; (3) Conflict Monitoring/Resolution; (4) Inhibitory control; and (5) Sustained Attention (for a more detailed description of the training program, see (Author)).

Genotyping Procedure and Classification

Variation in genes involved in the regulation of dopamine (COMT, DAT1, DRD4), serotonin (5HTT), and norepinephrine (MAOA3), which have been shown to interact with environmental factors influencing cognition and behavior during development, were included in the analysis. In addition, we included several genetic variations related to the modulation of acetylcholine (CHRNA4), and other brain factors (SNAP, DBH), which have been linked to ADHD and to individual differences in attention and other executive processes, to explore their potential interaction with the cognitive training program implemented here (Gosso et al., 2006; Parasuraman, Greenwood, Kumar, & Fossella, 2005).

DNA was isolated from saliva samples using Oragene collection kits (DNA Genotek Inc., Ontario, Canada) according to the manufacturer's instructions. Approximately, 10–40 ng of template was included in each PCR amplification, reactions contained 0.2 mM each deoxynucleotide, 0.2 μM each oligonucleotide, 0.05 U/μl recombinant Taq DNA polymerase with its 1× reaction buffer $(NH_4)_2SO_4$ (Thermo Fisher Scientific, Pittsburg, PA), and 8% (COMT, DAT1)—20%(DRD4) QuickExtract buffer V1.0 (Epicenter Biotechnologies, Madison, WI) in addition to PCR- specific optimizations. The DAT1 amplification contained 1.5 mM $MgCl_2$, 0.6 M betaine, and the oligonucleotide DAT1F 5′ - TGTGGTGTAGGGAACGGCCTGAG and DAT1R 5′-CTTCCTGGAGGTCACGGCTCAAGG (Shinohara et al., 2004). Amplification conditions were the following: 95°C 4 min; 35 × 94°C 30 s, 65°C 1 min, 72°C 30 s; 72°C 3 min. Amplified products were size separated on a 2% agarose gel (GenePure LE, BioExpress, Kaysville, UT) and visualized using ethidium bromide.

The COMT haplotype combinations as well as the different allele groups included in the analysis were arranged based on previous literature regarding their association with brain function or cognitive and behavioral performance. Accordingly, a 40 base pairs (bp) variable number tandem repeat (VNTR) located in the 3′ untranslated region (UTR) of the DAT1 gene was included in the study. The most common alleles of this polymorphism are the 9-r allele (a 440 bp product) and the 10r allele (a 480 bp 10-repeat product) (Kang, Palmatier, & Kidd, 1999). Children were grouped according to the presence (9/9 or 9/10) or absence (10/10) of the 9-r allele.

Variations in the gene encoding the catechol-*O*-methyltransferase (COMT) were included in the study considering both the SNP in codon 158 ($Val^{158}Met$) and the COMT haplotype. Compared to the COMT $Val^{158}Met$ genotype that reflects variation in enzyme activity, the COMT haplotypes have shown to reflect differences both at the enzyme expression level and the activity level. Diatchenko et al. (2005) identified three common COMT haplotypes that constitute a more sensitive measure of COMT expression levels and allows to categorize alleles into low, medium, and high levels of COMT enzymatic activity. In the present study, the COMT $Val^{158}Met$ genotype was grouped according to the presence or absence of the methionine allele (Met- Ab/Pr), and the COMT haplotype was amplified and identified as in Voelker et al. (2009). Three different groups were constituted. The LPS group, associated to higher levels of COMT enzymatic activity,

comprised children homozygous for the LPS allele as well as carriers of the LPS/APS genotype; the MSP group, associated to medium levels of COMT activity, were either homozygous for the APS allele or carriers of the HPS/LPS. Finally, children homozygous for the HPS allele and carriers of the APS/HPS genotype were included in the HPS group, a variation related to lower levels of COMT activity and higher levels of catecholamine-mediated neurotransmission (Diatchenko et al., 2006).

The DRD4 genotype was determined as previously published (Sheese, Rothbart, Voelker, & Posner, 2012). Individuals were grouped by the presence or absence of the 7-repeat allele (619 bp). The 5HTT human serotonin transporter genotype was grouped according to the presence or absence of the long (L- Ab/Pr) and short (S- Ab/Pr) variants of the polymorphism. Regarding the monoamine oxidase A, MAOA gene, a VNTR polymorphism within the promoter region was analyzed. Individuals were grouped according to the presence or absence of the 3-repeat allele. The CHRNA4 gene, also known as Human Neural Nicotinic Acetylcholine Receptor, subunit α4 was analyzed including the presence/absence of the *C* allele. For the synaptosomal-associated protein 25 or SNAP-25 gene, two different Single Nucleotide Polymorphisms (SNPs) were genotyped: the *Mnl*I and the *Ddel*I. The three most common variations within each SNP were included in the analysis: T–T, T–G, and G–G for the *Mnl*I and the T–C, T–T, and CC for the *Ddel*I SNP. Finally, variations within the Dopamine B-hydroxylase (DBH) gene, which is a G to A polymorphism, were included in the analysis grouped in relation to the presence or absence of the A allele. Table 1 shows the distribution of the sample regarding the genetic variations considered in the present study.

Cognitive Assessment Tasks

The Kaufman Brief Intelligence Test (K-BIT; (Kaufman & Kaufman, 2014)) was applied to assess children's Gf. Administration of the test takes approximately 20 min per child of preschool age. The Working Memory Span Subtest of the WISC-III, backwards condition (Wechsler, 1991), was used to measure processes of maintenance and manipulation of information.

Analysis Procedure

There are several machine learning methods that show similar performance as predictive classifiers over a wide range of applications, such as logistic regression, random forest, decision trees, and support vector machines, among others (e.g., Maroco et al., 2011). However, it has been found that in several conditions, ANN consistently outperform the other methods. ANN are very flexible when the objective is to maximize the precision in the prediction, and particularly effective when the data consist of variables with complex intercorrelations (Caruana & Niculescu-Mizil, 2006; Duin, 1996) (see King, Feng, & Sutherland, 1995 for a comparison among different machine learning methods). Specifically, this study used a multilayer perceptron network with a backpropagation algorithm. It is composed of nonlinear units which compute their activation level by summing all the weighted activations they receive; then, they transform their activation into a response via a nonlinear transfer function which establishes a relationship between the inputs and the weights they are assigned (see (David Garson, 1998) and Figure 1). In addition, a logistic regression analysis was used to analyze the same data in order to compare the predictive classification results of a more traditional approach with the ANN method. SPSS v.19 was used for the development and analysis of all ANN and LR predictive models.

Traditional measures were calculated including the determination of actual values and rates for true positive (TP), true negative (TN), false positive (FP), and false negative (FN) outcomes. In these analyses, both precision and recall or sensitivity, as outcome measures of the network, were given equal weight when evaluating the quality of the neural network results. Recall or sensitivity represents the proportion of correctly identified targets, out of all targets presented in the set, and is represented as: Recall = TP/(TP + FN). Precision represents the proportion of correctly identified targets, out of all identified targets by the system, and is represented as: Precision = TP/(TP + FP). Specificity is defined as the proportion of correctly rejected targets from all the targets that should have been rejected by the system, and it is expressed as Specificity = TN/(TN + FP). Other quality measure as F1-Score was calculated taking into account both false positives and false negatives and it is the harmonic mean of Precision and Recall.

Architecture of the Neural Networks

Two different neural networks (ANN) were developed as predictive systems for the classification of the pre- to post-training gains, one for each of the two standardized cognitive measures: Gf, on the one hand, and Working Memory Span (WM), on the other hand. ANN1 was developed to maximize the predictive classification of the "Zero or less" ("No-Gain" condition) and "Higher than zero" ("Gain" condition) in IQ-f change. Similarly, ANN2 was developed for the classification of "Zero or less" ("No-Gain" condition) and "Higher than zero" ("Gain" condition) in WM span change.

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. Literature suggests that about 2/3 of the cases in the available data set can be used for the

TABLE 1
Sample distribution by genetic variations

	Genetic variant							*Valid* N	*Missing*
	Haplotype group								
	LPS		MPS		HPS				
Frequency	35		16		7			58	8
Proportion	60.3		27.6		12.1				
	DAT1 10r Ab/Pr								
	10r - Pr		10r - Ab						
Frequency	50		9					59	7
Proportion	84.7		15.3						
	COMT Met-Ab/Pr								
	Met - Pr		Met - Ab						
Frequency	40		22					62	4
Proportion	64.5		35.5						
	DRD4 Genotype								
	2/3	2/4	4/4	4/5	4/7	4/8	7/7		
Frequency	1	4	33	4	17	2	1	62	4
Proportion	1.6	6.5	53.2	6.5	27.4	3.2	1.6		
	DRD4 7r Ab/Pr								
	7r - Pr		7r - Ab						
Frequency	18		44					62	4
Proportion	29.0		71.0						
	MAOA 3r Ab/Pr								
	3r - Pr		3r - Ab						
Frequency	43		19					62	4
Proportion	69.4		30.6						
	SNAP-25 MnlI polymorphism Genotype								
	G/G		T/G		T/T				
Frequency	10		31		21			62	4
Proportion	16.1		50.0		33.9				
	SNAP-25 DdelI polymorphism Genotype								
	C/C		T/C		T/T				
Frequency	1		17		44			62	4
Proportion	1.6		27.4		71.0				
	5HTT Genotype								
	S/S		S/L		L/L				
Frequency	17		28		17			62	4
Proportion	27.4		45.2		27.4				

TABLE 1
Continued

	Genetic variant		*Valid* N	*Missing*
	5HTT S-Ab/Pr			
	S - Pr	S - Ab		
Frequency	45	17	62	4
Proportion	72.6	27.4		
	5HTT L-Ab/Pr			
	L - Pr	L - Ab		
Frequency	45	17	62	4
Proportion	72.6	27.4		
	CHRNA4 C-Ab/Pr			
	C - Pr	C - Ab		
Frequency	42	20	62	4
Proportion	67.7	32.3		
	DBH A-Ab/Pr			
	A - Pr	A - Ab		
Frequency	47	15	62	4
Proportion	75.8	24.2		

Note: Each section presents the frequency, proportion, valid number of samples, and missing genetic classification for each genetic variant included in the analysis. The Ab/Pr categorization refers to the Absence or Presence of the genetic variant of interest for each gene included in the analysis.
Abbreviations: HPS, High Pain Sensitivity; LPS, Low Pain Sensitivity; MPS, Moderate Pain Sensitivity.

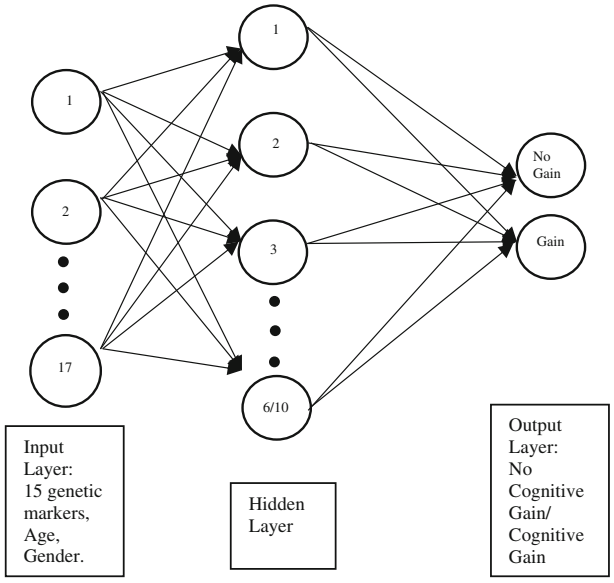

Fig. 1. Multilayer perceptron network architecture with one hidden layer used in this study.

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 of information on the input variables for each case). The remaining 1/3 of the data is used for the testing phase of the network.

The implementation and evaluation of the models was based on a systematic procedure suggested by (Rodríguez-Hernández, Musso, Kyndt, & Cascallar, 2021). During the training phase, the system evaluates the effect of the weight patterns on the precision of their classification of outputs, and then, through backpropagation, it adjusts those weights in a recursive fashion until they maximize the precision of the resulting classifications. During this training phase several models were attempted, and several modifications of the neural network parameters and hyperparamenters were explored, such as: learning persistence (controlling the continuation of training after no significant change in weights), 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 overfitting 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. The systematic changes to several parameters (i.e., five learning rate values by nine momentum values by two activation functions) led to 90 models for training and testing ANNs in each cognitive gain group. Finally, the model that achieved the best accuracy for both cognitive and noncognitive gain on

TABLE 2
Architecture of ANNs

Topology	*NNs fluid intelligence gain*	*NNs WM gain*
Initial learning rate values (×5)	0.4, 0.3, 0.2, 0.2, 0.1	0.4, 0.3, 0.2, 0.2, 0.1
Momentum values (×9)	Going from 0.1 to 0.9	Going from 0.1 to 0.9
Transfer function of the hidden layer	Hyperbolic tangent	Hyperbolic tangent
Transfer function of the output layer (×2)	Sigmoid and Softmax	Sigmoid and Softmax
Partitions data set	Training: 70% Testing: 30%	Training: 70% Testing: 30%
	Final NN fluid intelligence gain	Final NN WM gain
Training set data	78.8%	77.4%
Testing set data	21.2%	22.6%
Cross-entropy error	3.804	2.015
Stopping error	Two consecutive steps with no decrease in error	Two consecutive steps with no decrease in error
Number of input nodes	49	49
Number of output units	2 units: "Zero or less" "High than zero"	2 units: "Zero or less" "High than zero"
Number of hidden layers	One hidden layer with *six* units	One hidden layer with 10 units
Number of epochs for training	Automatically computed by the system	Automatically computed by the system
Method for rescaling covariates	Standardized method	Standardized method
Activation function for hidden layers	Hyperbolic tangent	Hyperbolic tangent
Activation and error function for output layer	Softmax. cross-entropy.	Softmax. cross-entropy
Methodology in the training phase	Online (one case by cycle)	Online
Parameters	Initial learning rate = 0.4	Initial learning rate = 0.4
	Momentum = 0.9	Momentum = 0.9
	Optimization algorithm: gradient descent	Optimization algorithm: gradient descent
	Minimum relative change in training error = 0.0001.	Minimum relative change in training error = 0.0001.

Note: The gradient descent optimization algorithm takes steps proportional to the negative of the approximate gradient of the function at the current point. Cross-entropy function accelerates the backpropagation algorithm, and it provides good overall network performance with relatively short stagnation periods.

the testing phase was selected for each one of the target classifications.

Each ANN introduced 17 input predictors: 15 specific genetic markers, gender, and age. They were used for the development of the vector–matrix containing all predictor variables for each student. Table 2 shows the topology of the explored ANNs and the architecture for each final model predicting gains in Gf and WM.

RESULTS

In order to evaluate the quality of the solutions achieved by each ANN, several measures were calculated. These results are presented in the Table 3. Accuracy measures refer to the percentages of the correct classifications in each group. As Table 3 shows, both ANN models achieved high accuracy for each of the outputs. In addition, the solutions have good "recall" (or sensitivity): the proportion of correctly identified targets, out of all targets actually presented in the set. The "precision" and "specificity" results were very good. The area under the ROC curve represents the true-positive rate (Sensitivity) plotted as a function of the false-positive rate (100 - Specificity) for different cut-off points and it can be viewed as a measure of the overall model performance across all possible thresholds, that is, how well it distinguishes between two groups.

ANN 1: Zero or Less Versus Higher than Zero Gain in Fluid Intelligence

Table 4 shows the "confusion matrix" representing all four outcomes for the training and testing phases.

Table 5 and Figure 2 show the actual predictive weight (importance for the classification) of the participating predictor variables (factors and covariates), as well as their normalized importance (expressed as percentages of the best predictor participation). COMT haplotype, age, gender and the presence or absence of 10r allele of the dopamine transporter (DAT1) gene were the top four predictors with the most significant importance in modeling Gf gains after training.

ANN 2: Zero or Less Versus Higher than Zero Gain in WM Span

Table 6 shows the "confusion matrix" representing all four outcomes for the training and testing phases.

TABLE 3
Measures for ANN and LR in the prediction of gain in fluid intelligence (Gf) and WM in the training and testing phases

	Fluid intelligence			*WM*		
	Artificial neural networks			*Artificial neural networks*		
Measures	*Training*	*Testing*	*Logistic regression*	*Training*	*Testing*	*Logistic regression*
Accuracy for "Higher than zero" group (TP)	96.3%	83.3%	82.6%	73.3%	80%	67.7%
Accuracy for "Zero or less" group (TN)	92.9%	100%	75.7%	96.2%	100%	88.9%
Overall accuracy (TP + TN)/(TP + FP + FN + TN)	95.1%	90.9%	79.5%	87.8%	91.7%	81.2%
Precision = TP/(TP + FP)	0.96	1	0.80	0.92	1	0.78
Sensitivity/Recall = TP/(TP + FN)	0.96	0.83	0.83	0.73	0.80	0.68
Specificity = TN/(TN + FP)	0.93	1	0.75	0.96	1	0.89
F1 Score (harmonic mean of PPV & TP 2TP/(2TP + FP + FN)	0.96	0.91	0.82	0.81	0.88	0.72
Area under the curve	–	0.971	–	–	0.953	–

Note: Precision represents the proportion of correctly identified targets. Out of all true targets presented to the system. Specificity is the proportion of correctly identified nontargets. Out of all true-nontargets presented in the set. The F1-Score is the harmonic mean of Precision and Recall taking both false positives and false negatives into account.
Abbreviations: FN, False Negatives; FP, False Positives; NPV, Negative Predicted Value; PPV, Positive Predicted Value; TN, True Negatives; TP, True Positives.

TABLE 4
Predictive classification of gains in Gf

		Classification		
		Predicted (%)		
Sample		*Zero or less*	*Higher than zero*	*Overall percent correct*
Training	Zero or less	92.9	7.1	
	higher than zero	3.7	96.3	
	Overall percent	34.1	65.9	
Testing	Zero or less	100	–	
	higher than zero	16.7	83.3	
	Overall percent	54.5	45.5	90.9

Table 7 and Figure 3 show the actual predictive weight (importance) of the participating predictor variables (factors and covariates), as well as their normalized importance. The COMT haplotype, age, presence, or absence of the 3r allele of the monoamine oxidase A (MAOA), and the DRD4 genotype were particularly informative in the modeling of the pre- to post-training change in WM.

Logistic Regression Results

A likelihood-ratio logistic regression (enter method) was carried out to predict children belonging to each cognitive gain group (Zero or less vs. higher than zero gain in Gf and WM) using the same predictors used with the ANN.

For the Gf gain model, the 17 predictors accounted for 46% (Nagelkerke $R^2 = 0.465$) of the variance in Gain/No-Gain group membership, but only gender reached significance ($B = 1.884$; $p < .05$). The Homer-Lemeshow test yielded a $X^{2\,(8)}$ of 4.933 ($p > .05$) suggesting the data had an adequate overall fit to the model. Accuracy and other quality measures are presented in the ANN table (Table 3) to facilitate the comparison.

For the prediction of WM gain, the 17 predictors accounted for 47% (Nagelkerke $R^2 = 0.472$) of the variance in Gain/No-Gain group membership, but no variable reached significance. The Homer-Lemeshow test yielded a $X^{2\,(7)}$ of 7.667 ($p > .05$) suggesting the data had an adequate overall fit to the model. Accuracy and other quality measures are presented in the ANN table (Table 3) to facilitate the comparison.

Differences between Predicted Gain Groups

We carried out a series of X^2 tests for categorical variables (genetic markers) and independent-samples t-test for continuous dependent variables (age in months) to test the differences (cognitive and genetic predictors) between Gain and No-Gain children groups predicted by the neural networks.

Gf gains and Group Differences

No significant differences were found among the three levels of Haplotype regarding the proportions of each Gf gain groups ($X^{2\,(2)} = 0.675$; $p > .05$).

TABLE 5
Relative importance and predictive weights of the variables participating in the model for the predictive classification of the gains in Gf

Predictor	*Importance*	*Predictive weights (%)*
COMT Haplotype	0.202	20.22
Age	0.122	12.16
Gender	0.110	11
DAT1 10r Ab/Pr	0.097	9.69
DRD4 Genotype	0.082	8.19
SNAP-25 *Mnl*I	0.074	7.39
DBH *A*- Ab/Pr	0.065	6.51
5HTT Genotype	0.063	6.29
5HTT *S*- Ab/Pr	0.043	4.33
MAOA *3r* Ab/Pr	0.038	3.79
COMT Val^{158}Met *Met*- Ab/Pr	0.028	2.81
CHRNA4 *C*- Ab/Pr	0.025	2.46
5HTT *L*- Ab/Pr	0.022	2.21
DRD4 *7*- Ab/Pr	0.017	1.72
SNAP-25 *Dde*I	0.012	1.22

Gender was associated with Gf gains ($X^{2\,(1)} = 4.880$; $p < .05$, *Cramer's* $V = 0.301$; $1 - \beta = 0.68$), 70.6% of the Gain group were males, while 60% of No-Gain group were females.

DAT1 10r Ab/Pr was related with Gf gains ($X^{2\,(1)} = 5.804$; Fisher's Exact Test $p < .05$; *Cramer's* $V = 0.328$; $1 - \beta = 0.74$). Most of the children in the Gain group were carriers of at least one copy of 10r allele (94.1%) while 70% of the No-Gain group were carriers of this allele. In addition, the DRD4 genotype was associated with Gf gains ($X^{2\,(5)} = 11.342$; $p < .05$; *Cramer's* $V = 0.458$; $1 - \beta = 0.83$), with 58.8% of cognitive Gain group carrying the 4/4 combination of the gene. No significant differences were found between groups for the other genetic markers.

Independent-samples t-test revealed that there was a significant difference between predicted Gain groups in Gf when considering the pre-test scores ($t^{(51)} = 3.013$; $p < .01$; *Cohen's* $d = 0.854$). Children who had benefited from the training program (Gain condition) were those who had lower pre-test Gf ($M = 100.18$; $SD = 16.576$) compared with children who had not benefited from the intervention (No-Gain condition $M = 109.95$; $SD = 10.94$; $1 - \beta = 1.00$). (See Table 8).

WM Gains and Group Differences

The COMT Haplotype was significantly associated with Gain and No-Gain differences in the backward span measure ($X^2\,(2) = 8.93$; $p < .01$; *Cramer's* $V = 0.392$). 72.2% of the No-Gain group belonged to the LPS group of the Haplotype, while 13.9% belonged to the MPS group, and the remaining 13.9% belonged to the HPS group. Regarding the Gain group, 50% of children belonged to the MPS group of the COMT Haplotype, 40.9% were part of the LPS group, and only the 9.1% belonged to the HPS group.

No significant differences were found between Gain and No-Gain groups for the rest of the genetic markers nor for gender. In addition, no significant differences were found between predicted Gain groups in the baseline pre-test value of Gf, WM, nor in age (see Table 9).

DISCUSSION

The strengths and advantages of the machine learning approach explain the high precision and accuracy findings in this study. One strength of this approach is to dispense with the assumptions required by traditional statistical predictive models (e.g., ordinary least squares regression, logistic regression) (Cascallar et al., 2014). Both ANN models were able to model nonlinear, complex, and not well-known relationships among several variables, in this case genetic factors, age and gender, while classifying individuals between cognitive Gain and No-Gain groups. The iterative process of error reduction during the training of the network, through the back-propagation algorithm, enables the ANN to "learn" the patterns associated with the outcome of each individual case in the training set. It then applies the model generated to the new data in the testing sample, generalizing the results.

An additional advantage of ANN is that it can use information from all the multiple interactions between all the predictors. While statistical significance is determined by the probability of error (p-value) in the estimate of the difference between values of a variable in different groups, the weights in an ANN are partially analogous to the coefficients in a general linear model (GLM). However, while in a GLM the weight in a model prediction represents the relative importance of a variable in its association with the dependent or outcome variable, in ANN there are many weights which connect one predictor to the outcome, including the weights influenced by the interaction effects

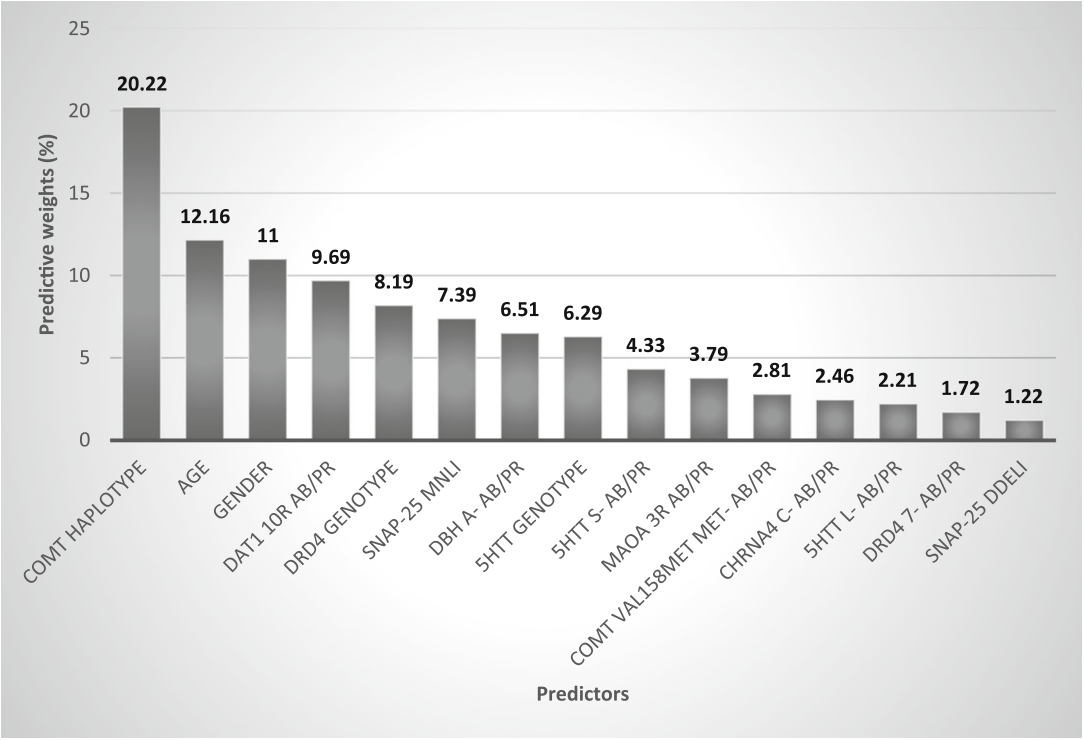

Fig. 2. Predictive weights of the variables participating in the model for the predictive classification of the gains in fluid intelligence (Gf).

TABLE 6
Predictive classification of gains in working memory (WM)

Classification				
		Predicted (%)		
Sample		*Zero or less*	*Higher than zero*	*Overall percent correct*
Training	Zero or less	96.2	3.8	
	higher than zero	26.7	73.3	
	Overall percent	70.7	29.3	
Testing	Zero or less	100	–	
	higher than zero	20	80	
	Overall percent	66.7	33.3	91.7

between all predictors, which are reflected in the hidden layers. Although this architecture makes an ANN very flexible and effective in modeling nonlinear conditions, it does present some difficulties for a correct interpretation. We follow (Garson, 2016) recommendation, expanded by (Goh, 1995) of considering all model weights, pooling and scaling all weights specific to a predictor, thus generating a single value (with a range from 0 to 1) that reflects the relative importance of the predictor on the corresponding outcome. It can then be considered that this predictive weight assigned to each predictor captures the complexity of the relationships in a manner that can only be achieved with an ANN. In any given comparison between groups with a significance test, we would only be capturing as significant those differences that, given the observed difference in the variable, and the size of the sample, would be below a certain probability of error. The ANN, on the other hand, informs on the effect size taking into account the full complexity of the model. Its reliability is measured by the various statistics used to evaluate the ANN results (e.g., F_1, precision, recall, sensitivity, etc.). Using all this information the ANN models were able to detect patterns among input predictors for each of the cases associated with a certain outcome. This is compatible with previous evidence regarding the robustness of neural models in the statistical sense, even when they are faced with a small number of data points (David Garson, 1998).

Classical statistical analyses revealed that children with lower Gf scores at baseline improved more after cognitive training than those starting with higher Gf. This result is in agreement with some previous findings (Diamond & Lee, 2011; Jaeggi et al., 2008; Mackey, Hill, Stone, & Bunge, 2011). However, in this study there are no baseline differences in WM span between Gain and No-Gain groups in Gf, contrary to the findings reported in previous studies (Foster, 2017; Fuchs et al., 2013). This result could be explained by the different WM tasks used: backward task span involves a simpler task compared to the complex span tasks used in those studies (Kane et al., 2005). The complex

TABLE 7
Relative importance and predictive weights of the variables participating in the model for the predictive classification of the gains in WM

Predictor	*Importance*	*Predictive weights (%)*
COMT Haplotype	0.284	28.38
Age	0.116	11.61
MAOA *3r* Ab/Pr	0.106	10.56
DRD4 Genotype	0.097	9.68
SNAP-25 *DdeI*	0.079	7.91
CHRNA4 *C-* Ab/Pr	0.077	7.72
SNAP-25 *MnlI*	0.044	4.41
5HTT Genotype	0.041	4.11
DRD4 *7* Ab/Pr	0.032	3.20
DBH *A-* Ab/Pr	0.026	2.60
COMT Val^{158}Met *Met-* Ab/Pr	0.026	2.58
5HTT *S-* Ab/Pr	0.023	2.31
DAT1 10r Ab/Pr	0.019	1.89
Gender	0.016	1.62
5HTT *L-* Ab/Pr	0.014	1.42

WM tasks present a secondary processing task that depletes capacity from the cognitive system, interfering with the storage of information, and they are more strongly correlated with Gf than memory span tasks (Kane et al., 2005).

In terms of the relationship between genetic variation and the prediction of cognitive gains given the training program implemented, both networks were able to achieve high predictive accuracy based on specific genetic information, gender, and age, as it was hypothesized. This could be explained because the selection of the predictors of the ANN was based on solid theoretical knowledge about which specific genetic markers are related with individual differences in brain activation, cognitive and behavioral outcomes, and susceptibility to environmental factors, as it is suggested by a structured neural network approach for modeling (Lee & Garver, 2005).

The results reported in this study suggest that particular variations in genes involved in the regulation of dopamine neurotransmission in the brain, among other individual differences as gender and age, affect children's susceptibility to benefit from executive attention training, a pattern that is consistent with previous studies (Brehmer et al., 2009; Söderqvist, 2012).

The machine learning approach used in this study facilitates the analysis of the differences in the contribution of a set of genes involved in cognitive functions that are highly correlated (Conway, Kane, & Engle, 2003; Kane et al., 2005). It was found that the COMT haplotype, age, gender, and the dopamine transporter DAT1 genotype were the top four predictors with the most significant importance in modeling Gf change after cognitive training. It was also determined that the COMT haplotype, age, the monoamine oxidase A (MAOA) genotype and the DRD4 genotype were particularly informative in the modeling of the pre- to post-training change in WM.

The results achieved in this study, given the statistical robustness offered by machine learning analysis, draw a pattern where multiple genetic markers, along with other constitutional factors such as gender or age, contribute to children's susceptibility to benefit from a systematic cognitive training program. This pattern relates to the concept known as "plasticity genes" introduced by Belsky et al. (2009) to represent how some genetic markers operate to make individuals more susceptible to the influence of particular environmental factors "for better and for worse" (Belsky et al., 2009). An example of this was reported by Belsky & Beaver (2011), who aimed at understanding the cumulative-genetic plasticity showed by dopamine-related genetic markers including alleles of the DAT1, the DRD4 and the MAOA genes, in relation to the influence of parenting on adolescent's self-regulation skills. In line with our results, (Belsky & Beaver, 2011) demonstrated that both gender and dopamine-genetic markers interacted with environmental variables associated to parenting supportiveness to explain adolescents' individual differences in the development of self-regulation.

It is important to note that, although the pattern of contribution differs between WM and Gf gains, there is an important commonality among the genetic markers found to be predictive of children's susceptibility to cognitive training in both cases: they are all involved in the regulation of dopamine levels in brain areas that are highly associated to the performance of executive functions and other cognitive processes typically recruited by Gf tasks (Duncan & Owen, 2000). Previous studies have reported that cognitive training is associated to increased BOLD activity (Dahlin, Nyberg, Bäckman, & Neely, 2008) and DA release (Bäckman & Nyberg, 2013) in the same areas that are involved in the performance of the cognitive tasks that are subject to training. Therefore, efficiency of dopaminergic pathways does appear to play an important role defining whether an individual can benefit from a cognitive training program and to this extent, Gene × Training studies are key to understand the mechanisms that underlie such relationship.

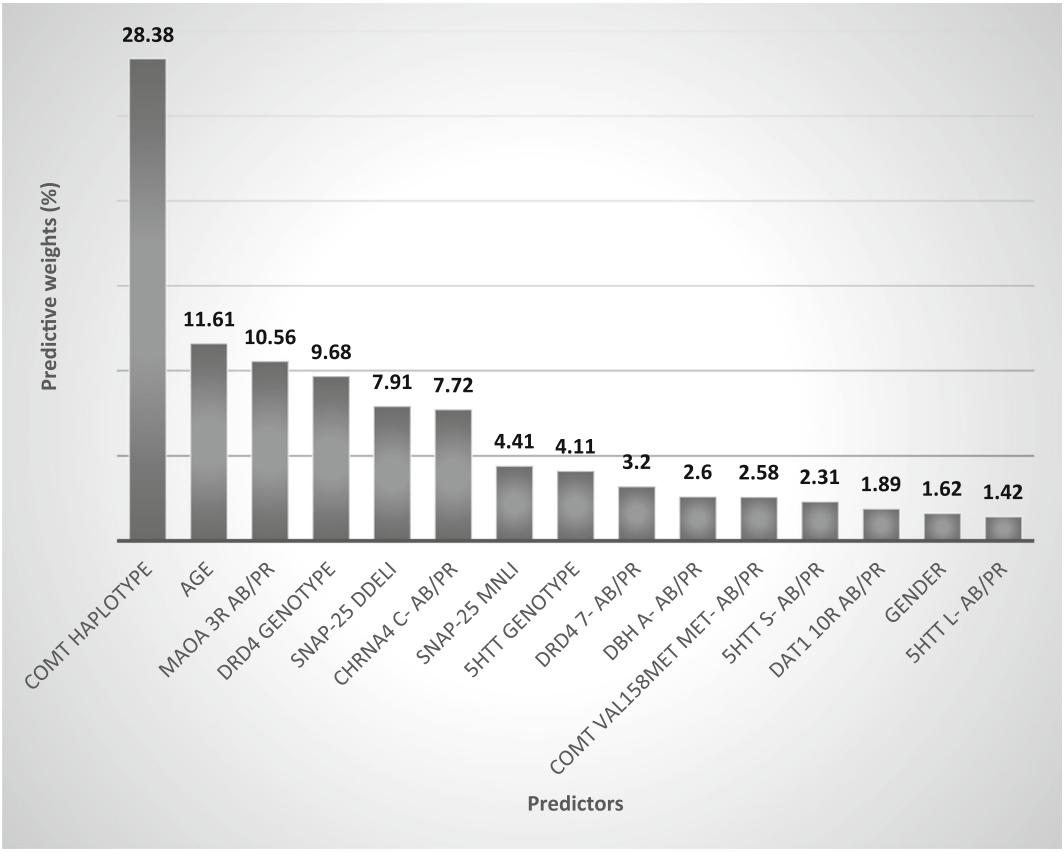

Fig. 3. Predictive weights of the variables participating in the model for the predictive classification of the gains in working memory (WM).

TABLE 8
Mean differences between Gf gain and No-Gain predicted groups in pre-testing cognitive measures and age

Cognitive measures pre-test	*No-Gain group mean (SD)*	*Gain group mean (SD)*	t-*test (df)*	p
Backward span	1.70 (1.42)	2.26 (1.26)	−1.516 (52)	.135
Fluid intelligence	109.95 (10.94)	100.18 (11.72)	3.013 (51)	.004*
Age	61.08 (7.05)	64.25 (7.77)	−1.496 (52)	.141

*The difference was significant at level $p < .01$.

TABLE 9
Mean differences between WM gain and No-Gain predicted groups in pre-testing cognitive measures and age

Cognitive measures pre-test	*No-Gain group mean (SD)*	*Gain group mean (SD)*	*t- test (df)*	p
Backward span	2.14 (1.417)	1.75 (1.065)	0.973 (51)	.335
Fluid intelligence	103.89 (12.405)	103.80 (12.542)	0.024 (50)	.981
Age	63.45 (7.514)	61.07 (7.766)	1.048 (51)	.30

Evidence involving humans and other species have shown that there is an inverted-U shape relationship between prefrontal dopaminergic function and performance of executive function tasks (Barnett et al., 2009; Williams & Goldman-Rakic, 1995). Using traditional statistical analysis, several studies have shown that DA genetic variations that lead to a more efficient dopaminergic function, are also linked to increased cognitive gains after training. For instance, (Brehmer et al., 2009) found that individuals carrying the 9/10 genotype of the DAT1 gene, a variation associated to more active DA pathways within the striatum, could benefit to a greater extent from a computerized WM training program. This pattern has been also observed with the COMT gene in younger participants where carriers of the more efficient Met allele have been found to have increased WM plasticity (Zhao et al., 2020).

In the present study, the COMT haplotype showed to be the strongest predictor of children's capacity to benefit from training for both Gf and WM outcomes. Previous research has shown that, compared to the study of individual polymorphism within the catechol-O-methyltransferase (COMT) gene, the use of COMT haplotype provides a better understanding of the COMT enzyme function, reflecting

both its expression and its activity. In vitro studies have demonstrated that the LPS haplotype is related to a higher level of enzymatic activity (i.e., lower levels of dopamine transmission), compared to the APS or the HPS haplotypes (Diatchenko et al., 2005). Several studies have reported that the COMT haplotypes are related to children's cognitive performance in a variety of measures including WM and inhibition (Barnett et al., 2009) and they have been shown to interact with environmental factors to influence the development of attention in younger children (Voelker et al., 2009).

Given that the developmental changes observed in COMT activity within the prefrontal cortex follow the same inverted U-shape relationship between DA availability and cognitive function (Tunbridge et al., 2007), it can be hypothesized that efficiency of the COMT enzyme within the prefrontal cortex during childhood may underlie the individual differences that lead to an increased cognitive plasticity for some children.

Some limitations of this study have to be considered. First, the sample size for modeling ANN was rather small so we could not generalize the models on a different random set of data. We have to consider that this type of machine learning approach has been used very recently to predict developmental or training changes, so this study has an exploratory scope. Also, given the fact that only recently machine learning approaches have been used to predict developmental and training data, this study should be considered exploratory in this realm. As such, further research should use larger samples to validate the models obtained, thus diminishing the risk of overfitting the data and increasing the generalization of the networks' functioning. Second, this study involved only a few variables such as certain genetic variations, age, and gender. Future studies should have to go beyond unmodifiable individual differences, including temperament and environmental factors in the pre-training baseline.

Acknowledgments—We thank M. Rosario Rueda and her research team from the Department of Experimental Psychology and Center for Research on Mind, Brain and Behavior (CIMCYC), Universidad de Granada, for providing access to the data set used for the application of the machine learning approach. We are also grateful to the children and caregivers who participated in the study.

Conflict of interest

The authors declare no conflict of interest.

Ethics statement

Ethical approval was obtained from the Research Ethics Committee of the University of Granada. The treatment of subjects was in accordance with established ethical standards of the Ethical Principles of Psychologists and Code of Conduct (APA, 2010). Parents were informed about the purpose of the study, and they gave written consent to participate in the study.

REFERENCES

Alloway, T. P., & Alloway, R. G. (2010). Investigating the predictive roles of working memory and IQ in academic attainment. *Journal of Experimental Child Psychology*, *106*, 20–29. https://doi.org/10.1016/J.JECP.2009.11.003

Arteaga Díaz, G., & Pimienta Jiménez, H. (2006). Memoria operativa y circuitos corticales. *Revista de la Facultad de Medicina*, *54*, 248–268. http://www.scielo.org.co/scielo.php?pid=S0120-00112006000400005&script=sci_abstract

Bäckman, L., & Nyberg, L. (2013). Dopamine and training-related working-memory improvement. *Neuroscience and Biobehavioral Reviews*, *37*, 2209–2219. https://doi.org/10.1016/j.neubiorev.2013.01.014

Bakermans-Kranenburg, M. J., & Van Ijzendoorn, M. H. (2006). Gene-environment interaction of the dopamine D4 receptor (DRD4) and observed maternal insensitivity predicting externalizing behavior in preschoolers. *Developmental Psychobiology*, *48*, 406–409. https://doi.org/10.1002/dev.20152

Bansal, A., Kauffman, R. J., & Weitz, R. R. (1993). Comparing the performance of regression and neural networks as data quality varies: A business value approach, NYU Work. Paper No. IS-93-34. https://ssrn.com/abstract=1288429.

Barnett, J. H., Heron, J., Goldman, D., Jones, P. B., & Xu, K. (2009). Effects of catechol-O-methyltransferase on normal variation in the cognitive function of children. *The American journal of psychiatry*, *166*(8), 909–16.

Bellander, M., Bäckman, L., Liu, T., Schjeide, B. M. M., Bertram, L., Schmiedek, F., … Lövdén, M. (2015). Lower baseline performance but greater plasticity of working memory for carriers of the val allele of the COMT val 158 met polymorphism. *Neuropsychology*, *29*, 247–254. https://doi.org/10.1037/neu0000088

Belsky, J., & Beaver, K. M. (2011). Cumulative-genetic plasticity, parenting and adolescent self-regulation. *Journal of Child Psychology and Psychiatry*, *52*, 619–626. https://doi.org/10.1111/J.1469-7610.2010.02327.X

Belsky, J., Jonassaint, C., Pluess, M., Stanton, M., Brummett, B., & Williams, R. (2009). Vulnerability genes or plasticity genes? *Molecular Psychiatry*, *14*(8), 746–754. https://doi.org/10.1038/mp.2009.44

Brehmer, L. B. Y., Westerberg, H., Bellander, M., Fürth, D., & Karlsson, S. B. (2009). Working memory plasticity modulated by dopamine transporter genotype. *Neuroscience Letters*, *467*, 117–120. https://doi.org/10.1016/j.neulet.2009.10.018

Caruana, R., & Niculescu-Mizil, A. (2006). An empirical comparison metrics of supervised learning algorithms using different performance learning, in: *23rd Int. Conf. Mach.*, Pittsburgh, PA.

Cascallar, E., Boekaerts, M., & Costigan, T. (2006). Assessment in the Evaluation of Self-Regulation as a Process. *Educational Psychology Review*, *18*(3), 297–306.

Cascallar, E., Musso, M., Kyndt, E., & Dochy, F. (2014). Modelling for understanding AND for prediction/classification-the power of neural networks in research. *Frontline Learning Research*, *6*, 67–81. https://doi.org/10.14786/flr.v2i5.135

Caspi, A., Mcclay, J., Moffitt, T. E., Mill, J., Martin, J., Craig, I. W., … Poulton, R. (2002). Role of genotype in the cycle of violence in maltreated children. *Science*, *297*(5582), 851–854. http://doi.org/10.1126/science.1072290

Colzato, L. S., van den Wildenberg, W. P. M., & Hommel, B. (2014). Cognitive control and the COMT Val158Met polymorphism: Genetic modulation of videogame training and transfer to task-switching efficiency. *Psychological Research*, *78*, 670–678. https://doi.org/10.1007/s00426-013-0514-8

Conway, A. R. A., Kane, M. J., & Al, C. E. T. (2005). Working memory span tasks: A methodological review and user' s guide. *Psychonomic Bulletin & Review*, *12*, 769–786.

Conway, A. R. A., Kane, M. J., & Engle, R. W. (2003). Working memory capacity and its relation to general intelligence. *Trends in Cognitive Sciences*, *7*, 547–552. https://doi.org/10.1016/j.tics.2003.10.005

Dahlin, E., Nyberg, L., Bäckman, L., & Neely, A. S. (2008). Plasticity of executive functioning in young and older adults: Immediate training gains. *Transfer, and Long-Term Maintenance*, *23*, 720–730. https://doi.org/10.1037/a0014296

David Garson, G. (1998). Neural networks. *An Introductory Guide for Social Scientists*, *28*, 753. https://doi.org/10.2307/2655607

Detienne, K. B., Detienne, D. H., & Joshi, S. A. (2003). Neural networks as statistical tools for. *Business Research*, *6*, 236–265. https://doi.org/10.1177/1094428103251907

Diamond, A., & Lee, K. (2011). Interventions shown to aid executive function development in children 4–12 years old. *Science*, *333*, 959–964. https://doi.org/10.1126/science.1204529.

Diatchenko, L., Nackley, A. G., Slade, G. D., Bhalang, K., Belfer, I., Max, M. B., … Maixner, W. (2006). Catechol-O-methyltransferase gene polymorphisms are associated with multiple pain-evoking stimuli. *Pain*, *125*, 216–224.

Diatchenko, L., Romanov, S., Malinina, I., Clarke, J., Tchivilev, I., Li, X., & Makarov, S. S. (2005). Identification of novel mediators of NF-κB through genome-wide survey of monocyte adherence-induced genes. *Journal of Leukocyte Biology*, *78*, 1366–1377. https://doi.org/10.1189/JLB.0405211

Duin, R. P. W. (1996). A note on comparing classifiers. *Pattern Recognition Letters*, *17*, 529–536. https://doi.org/10.1016/0167-8655(95)00113-1

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*, 163–170. https://doi.org/10.1109/HICSS.1991.184056

Duncan, J., & Owen, A. M. (2000). Common regions of the human frontal lobe recruited by diverse cognitive demands. *Trends in Neurosciences*, *23*, 475–483.

Engle, R. W., & Kane, M. J. (2003). Executive attention, working memory capacity, and a two-factor theory of cognitive control. *Psychology of Learning and Motivation*, *44*, 145–199. https://doi.org/10.1016/S0079-7421(03)44005-X

Engle, R. W., Laughlin, J. E., Tuholski, S. W., & Conway, A. R. A. (1999). Working memory, short-term memory, and general fluid intelligence: A latent-variable approach. *Journal of Experimental Psychology. General*, *128*, 309–331. https://doi.org/10.1037/0096-3445.128.3.309

Everson, H. T., Chance, D., & Lykins, S. (1994). Using Artificial Neural Networks in Educational Research: Some Comparisons with Linear Statistical Models. Paper presented at the *Annual Meeting of the National Council on Measurement in Education* (New Orleans, LA, April 5-7, 1994)

Everson, H. T. (1995). Modeling the student in intelligent tutoring sytems: The promise of a new psychometrics. *Instructional Science*, *23*(5), 433–452. https://doi.org/10.1007/BF00896881

Fossella, J., Posner, M. I., Fan, J., Swanson, J. M., & Pfaff, D. W. (2002). Attentional phenotypes for the analysis of higher mental function. *Scientific World Journal*, *2*, 217–223. https://doi.org/10.1100/tsw.2002.93

Foster, J. L. (2017). Do the effects of working memory training depend on baseline ability level? *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *43*, 1677–1689. https://doi.org/10.1037/XLM0000426

Fuchs, L. S., Schumacher, R. F., Sterba, S. K., Long, J., Namkung, J., Malone, A., … Changas, P. (2013). Does working memory moderate the effects of fraction intervention? *An Aptitude-Treatment Interaction*, *106*, 499–514. https://doi.org/10.1037/a0034341

Garson, G. D. (2016). A comparison of neural network and expert systems algorithms with common multivariate procedures for analysis of social science data. *Social Science Computer Review*, *9*, 399–434. https://doi.org/10.1177/089443939100900304

Goh, A. T. C. (1995). Back-propagation neural networks for modeling complex systems. *Artificial Intelligence in Engineering*, *9*, 143–151. https://doi.org/10.1016/0954-1810(94)00011-S

Gorr, W. L. (1994). Editorial: Research prospective on neural network forecasting. *International Journal of Forecasting*, *10*, 1–4. https://doi.org/10.1016/0169-2070(94)90044-2

Gosso, M. F., De Geus, E. J. C., Van Belzen, M. J., Polderman, T. J. C., Heutink, P., Boomsma, D. I., & Posthuma, D. (2006). The SNAP-25 gene is associated with cognitive ability: Evidence from a family-based study in two independent Dutch cohorts. *Molecular Psychiatry*, *11*(9), 878–886. https://doi.org/10.1038/sj.mp.4001868

Hambrick, D. Z., Oswald, F. L., Darowski, E. S., Rench, T. A., & Brou, R. (2010). Predictors of multitasking performance in a synthetic work paradigm. *Applied Cognitive Psychology*, *24*, 1149–1167. https://doi.org/10.1002/ACP.1624

Hardgrave, B. C., Wilson, R. L., & Walstrom, K. A. (1994). Predicting graduate student success: A comparison of neural networks and traditional techniques. *Computers and Operations Research*, *21*, 249–263. https://doi.org/10.1016/0305-0548(94)90088-4

Heitz, R. P., Unsworth, N., & Engle, R. W. (2005). Working memory capacity, attention control, and fluid intelligence. In *Handbook of understanding and measuring intelligence.* (pp. 61–78). California, USA: SAGE Publications Inc. https://doi.org/10.4135/9781452233529.n5

Horn, J. L., & Cattell, R. B. (1966). Refinement and test of the theory of fluid and crystallized general intelligences. *Journal of Education & Psychology*, *57*, 253–270. https://doi.org/10.1037/H0023816

Jaeggi, S. M., Buschkuehl, M., Jonides, J., & Perrig, W. J. (2008). Improving fluid intelligence with training on working memory. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(19), 6829–6833. https://doi.org/10.1073/pnas.0801268105

Jaeggi, S. M., Buschkuehl, M., Jonides, J., & Shah, P. (2011). Short- and long-term benefits of cognitive training. *Proceedings of the National Academy of Sciences*, *108*, 10081–10086. https://doi.org/10.1073/pnas.1103228108

Kane, M. J., Conway, A. R. a., Hambrick, D. Z., & Engle, R. W. (2007). Variation in working memory capacity as variation in executive attention and control. In A. R. A. Conway, C. Jarrold, M. J. Kane, A. Miyake & J. N. Towse (Eds.), *Variation in Working Memory*, (pp. 21–46). New York: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195168648.003.0002

Kane, M. J., Hambrick, D. Z., Conway, A. R. A., Ackerman, L., Beier, M. E., & Boyle, M. O. (2005). Working Memory Capacity and Fluid Intelligence Are Strongly Related Constructs: Comment on Ackerman, Beier, and Boyle (2005). *Psychological Bulletin*, *131*, 66–71. https://doi.org/10.1037/0033-2909.131.1.66.

Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., Engle, R. W., … Wilke, A. (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. https://doi.org/10.1037/0096-3445.133.2.189

Kang, A. M., Palmatier, M. A., & Kidd, K. K. (1999). Global variation of a 40-bp VNTR in the 3′-untranslated region of the dopamine transporter gene (SLC6A3). *Biological Psychiatry*, *46*, 151–160. https://doi.org/10.1016/S0006-3223(99)00101-8

Karbach, J., & Kray, J. (2009). How useful is executive control training? Age differences in near and far transfer of task-switching training. *Developmental Science*, *12*, 978–990. https://doi.org/10.1111/j.1467-7687.2009.00846.x

Kaufman, A. S., & Kaufman, N. L. (2014). Kaufman brief intelligence test. In *Encyclopedia of Special Education*. (2nd ed.). Hoboken (NJ): John Wiley & Sons, Inc. https://doi.org/10.1002/9781118660584.ESE1325

Kerns, K. A., Eso, K., & Thomson, J. (1999). Investigation of a Direct Intervention for Improving Attention in Young Children with ADHD. *Developmental Neuropsychology*, *16*, 273–295. https://doi.org/10.1207/S15326942DN1602_9

King, R. D., Feng, C., & Sutherland, A. (1995). Statlog: Comparison of classification algorithms on large real-world problems. *Applied Artificial Intelligence*, *9*, 289–333. https://doi.org/10.1080/08839519508945477

Klingberg, T. (2010). Training and plasticity of working memory. *Trends in Cognitive Sciences*, *14*, 317–324. https://doi.org/10.1016/j.tics.2010.05.002

Kochanska, G., Philibert, R. A., & Barry, R. A. (2009). Interplay of genes and early mother–child relationship in the development of self-regulation from toddler to preschool age. *Journal of Child Psychology and Psychiatry*, *50*, 1331–1338. https://doi.org/10.1111/j.1469-7610.2008.02050.x

Lee, C., & Garver, M. S. (2005). Structured Neural Network Techniques for Modeling Loyalty and Profitability. *Proceedings of the Thirtieth Annual SAS Users Group International Conference*. Cary, NC: SAS Institute Inc.

Lynn, R., Meisenberg, G., Mikk, J., & Williams, A. (2007). National IQs predict differences in scholastic achievement in 67 countries. *Journal of Biosocial Science*, *39*, 861–874. https://doi.org/10.1017/S0021932007001964

Mackey, A. P., Hill, S. S., Stone, S. I., & Bunge, S. A. (2011). Differential effects of reasoning and speed training in children. *Developmental Science*, *14*, 582–590. https://doi.org/10.1111/J.1467-7687.2010.01005.X

Maroco, J., Silva, D., Rodrigues, A., Guerreiro, M., Santana, I. & de Mendonça, A. (2011) Data Mining Methods in the Prediction of Dementia: A Real-Data Comparison of the Accuracy, Sensitivity and Specificity of Linear Discriminant Analysis, Logistic Regression, Neural Networks, Support Vector Machines, Classification Trees and Random Forests. *BMC Research Notes*, *4*, 299. https://doi.org/10.1186/1756-0500-4-299

Melby-Lervåg, M., & Hulme, C. (2012). Is working memory training effective? A meta-analytic review. *Developmental Psychology*, *49*, 270–291. https://doi.org/10.1037/a0028228

Minear, M., & Shah, P. (2008). Training and transfer effects in task switching. *Memory and Cognition*, *36*, 1470–1483. https://doi.org/10.3758/MC.336.8.1470

Musso, M. F. (2016). Understanding the underpinnings of academic performance: The relationship of basic cognitive processes, self-regulation factors and learning strategies with task characteristics in the assessment and prediction of academic performance. *Doctoral Thesis*. KU Leuven, Belgium, undefined. https://www.mendeley.com/catalogue/2dc21e11-12f9-392b-950e-f1a4b8e1f5b6/?utm_source=desktop&utm_medium=1.19.8&utm_campaign=open_catalog&userDocumentId=%7Bc8267624-fa5e-366c-bf76-b02b169fbbd3%7D

Musso, M. F. & Cascallar, E. C. (2009). New Approaches for Improved Quality in Educational Assessments: Using Automated Predictive Systems in Reading and Mathematics. *Problems of Education in the 21st Century*, *17*, 134–151.

Musso, M. F., Cascallar, E. C., Bostani, N., & Crawford, M. (2020). Identifying Reliable Predictors of Educational Outcomes Through Machine-Learning Predictive Modeling. *Frontiers in Education*, *5*. https://doi.org/10.3389/feduc.2020.00104

Musso, M. F., Hernández, C. F. R., & Cascallar, E. C. (2020). Predicting key educational outcomes in academic trajectories: a machine-learning approach. *Higher Education*, *80*, 875–894. https://doi.org/10.1007/s10734-020-00520-7

Musso, M., Kyndt, E., Cascallar, E., & Dochy, F. (2012). Predicting Mathematical Performance: The Effect of Cognitive Processes and Self-Regulation Factors. *Education Research International*, *2012*, 1–13. https://doi.org/10.1155/2012/250719

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

Neal, W. D., & Wurst, J. (2001). Advances in market segmentation. *Marketing Research*, *13*(1), 14–19.

Neville, H. J., Stevens, C., Pakulak, E., Bell, T. A., Fanning, J., Klein, S., & Isbell, E. (2013). Family-based training program improves brain function, cognition, and behavior in lower socioeconomic status preschoolers. *Proceedings of the National Academy of Sciences of the United States of America*, *110*, 12138–12143. https://doi.org/10.1073/pnas.1304437110

Parasuraman, R., Greenwood, P. M., Kumar, R., & Fossella, J. (2005). Beyond heritability neurotransmitter genes differentially

modulate visuospatial attention and working memory. *Psychological Science*, *16*(3), 200. https://doi.org/10.1111/j.0956-7976.2005.00804.x

Petersen, S. E., & Posner, M. I. (2012). The attention system of the human brain: 20 years after. *Annual Review of Neuroscience*, *35*, 73–89. https://doi.org/10.1146/annurev-neuro-062111-150525

Posner, M. I., & Petersen, S. E. (1990). The attention system of the human brain. *Annual Review of Neuroscience*, *13*(1), 25–42.

Posner, M. I., Rothbart, M. K., & Sheese, B. E. (2007). Attention genes. *Developmental Science*, *10*, 24–29. https://doi.org/10.1111/J.1467-7687.2007.00559.X

Pozuelos, J. P., Combita, L. M., Abundis, A., Paz-Alonso, P. M., Conejero, Á., Guerra, S., & Rueda, M. R. (2019). Metacognitive scaffolding boosts cognitive and neural benefits following executive attention training in children. *Developmental Science*, *22*(2). https://doi.org/10.1111/desc.12756

Redick, T. S. (2015). Working memory training and interpreting interactions in intelligence interventions. *Intelligence*, *50*, 14–20. https://doi.org/10.1016/J.INTELL.2015.01.014

Redick, T. S., Melby-Lervåg, M., & Hulme, C. (2016). Corresponding authors: Working memory training does not improve performance on measures of intelligence or other measures of "far transfer": Evidence from a meta-analytic review. *Perspectives on Psychological Science*, *11*, 512–534. https://doi.org/10.1177/1745691616635612

Rodríguez-Hernández, C. F., Musso, M., Kyndt, E., & Cascallar, E. (2021). Artificial neural networks in academic performance prediction: Systematic implementation and predictor evaluation. *Computers and Education: Artificial Intelligence*, *2*, 100018. https://doi.org/10.1016/J.CAEAI.2021.100018

Rueda, M. R. (2018). Attention in the heart of intelligence. *Trends in Neuroscience and Education*, *13*, 26–33. https://doi.org/10.1016/j.tine.2018.11.003

Rueda, M. R., Checa, P., & Cómbita, L. M. (2012). Enhanced efficiency of the executive attention network after training in preschool children: immediate changes and effects after two months. *Developmental Cognitive Neuroscience*, *2* Suppl 1, S192–204. https://doi.org/10.1016/j.dcn.2011.09.004

Rueda, M. R., Checa, P., & Rothbart, M. K. (2010). Contributions of attentional control to socioemotional and academic development. *Early Education and Development*, *21*(5), 744–764. https://doi.org/10.1080/10409289.2010.510055

Rueda, M. R., Rothbart, M. K., McCandliss, B. D., Saccomanno, L., & Posner, M. I. (2005). Training, maturation, and genetic influences on the development of executive attention. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(41), 14931–14936. https://doi.org/10.1073/pnas.0506897102

Sheese, B. E., Rothbart, M. K., Voelker, P. M., & Posner, M. I. (2012). The dopamine receptor D4 gene 7-repeat allele interacts with parenting quality to predict effortful control in four-year-old children. *Child Development Research*, *2012*, 1–6. https://doi.org/10.1155/2012/863242

Sheese, B. E., Voelker, P. M., Rothbart, M. K., & Posner, M. I. (2007). Parenting quality interacts with genetic variation in dopamine receptor D4 to influence temperament in early childhood. *Development and Psychopathology*, *19*, 1039–1046. https://doi.org/10.1017/S0954579407000521

Shinohara, M., Mizushima, H., Hirano, M., Shioe, K., Nakazawa, M., Hiejima, Y., … Kanba, S. (2004). Eating disorders with binge-eating behaviour are associated with the s allele of the 3′-UTR VNTR polymorphism of the dopamine transporter gene. *J. Psychiatry Neurosci.* *29*, 134–137.

Shipstead, Z., Harrison, T. L., & Engle, R. W. (2016). Working memory capacity and fluid intelligence: Maintenance and disengagement. *Perspectives on Psychological Science*, *11*, 771–799. https://doi.org/10.1177/1745691616650647

Shipstead, Z., Redick, T. S., & Engle, R. W. (2012). Is working memory training effective? *Psychological Bulletin*, *138*, 628–654. https://doi.org/10.1037/A0027473

Söderqvist, S. (2012) *Probing and pushing potential genetics, development and training of cognitive functions.* (pp. 2012). Stockholm: Karolinska institutet.

Söderqvist, S., Bergman Nutley, S., Ottersen, J., Grill, K. M., & Klingberg, T. (2012). Computerized training of non-verbal reasoning and working memory in children with intellectual disability. *Frontiers in Human Neuroscience*, *6*, 271. https://doi.org/10.3389/FNHUM.2012.00271

Söderqvist, S., Nutley, S. B., Peyrard-Janvid, M., Matsson, H., Humphreys, K., Kere, J., & Klingberg, T. (2012). Dopamine, working memory, and training induced plasticity: Implications for developmental research. *Developmental Psychology*, *48*, 836–843. https://doi.org/10.1037/a0026179

Studer-Luethi, B., Bauer, C., & Perrig, W. J. (2016). Working memory training in children: Effectiveness depends on temperament. *Memory and Cognition*, *44*, 171–186. https://doi.org/10.3758/s13421-015-0548-9

Swanson, H. L., Moran, A., Lussier, C., & Fung, W. (2014). The effect of explicit and direct generative strategy training and working memory on word problem-solving accuracy in children at risk for math difficulties. *Learning Disability Quarterly*, *37*, 111–123. https://doi.org/10.1177/0731948713507264

Thorell, L. B., Lindqvist, S., Nutley, S. B., Bohlin, G., & Klingberg, T. (2009). Training and transfer effects of executive functions in preschool children. *Developmental Science*, *12*, 106–113. https://doi.org/10.1111/j.1467-7687.2008.00745.x

Tunbridge, E., Weickert, C., Kleinman, J., Herman, M., Chen, J., Kolachana, B., … Weinberger, D. (2007). Catechol-o-methyltransferase enzyme activity and protein expression in human prefrontal cortex across the postnatal lifespan. *Cerebral Cortex*, *17*, 1206–1212. https://doi.org/10.1093/CERCOR/BHL032

Unsworth, N., Redick, T. S., Heitz, R. P., Broadway, J. M., & Engle, R. W. (2009). Complex working memory span tasks and higher-order cognition: A latent-variable analysis of the relationship between processing and storage. *Memory*, *17*, 635–654. https://doi.org/10.1080/09658210902998047

Voelker, P., Sheese, B. E., Rothbart, M. K., & Posner, M. I. (2009). Variations in catechol-O-methyltransferase gene interact with parenting to influence attention in early development. *Neuroscience*, *164*, 121–130. https://doi.org/10.1016/j.neuroscience.2009.05.059

Wechsler, D. (1991) *Wechsler intelligence scale for children- third edition (WISC-III).* (3rd ed.). San Antonio, TX: Psychological Corporation.

White, H., & Racine, J. (2001). Statistical inference, the bootstrap, and neural-network modeling with application to foreign

exchange rates. *IEEE Transactions on Neural Networks*, *12*, 657–673. https://doi.org/10.1109/72.935080

Williams, G. V., & Goldman-Rakic, P. S. (1995). Modulation of memory fields by dopamine dl receptors in prefrontal cortex. *Nature*, *376*(6541), 572–575. https://doi.org/10.1038/376572a0

Zhao, W., Huang, L., Li, Y., Zhang, Q., Chen, X., Fu, W., … Li, J. (2020). Evidence for the contribution of COMT gene Val158/108Met polymorphism (rs4680) to working memory training-related prefrontal plasticity. *Brain and Behavior: A Cognitive Neuroscience Perspective*, *10*, e01523. https://doi.org/10.1002/brb3.1523