Trends in Neuroscience and Education 17 (2019) 100121

Contents lists available at ScienceDirect

Trends in Neuroscience and Education

journal homepage: www.elsevier.com/locate/tine

Response suppression, strategy application, and working memory in the prediction of academic performance and classroom misbehavior: A neuropsychological approach

Graham Pluck*, David Villagomez-Pacheco, María Isabel Karolys, María Emilia Montaño-Córdova, Pamela Almeida-Meza

Institute of Neurosciences, Universidad San Francisco de Quito, Diego de Robles y Vía Interoceánica, Cumbayá, Ecuador

ARTICLE INFO

Keywords:
Response suppression
Inhibition
Executive function
Frontal lobes
Academic achievement
Behavior problems

ABSTRACT

Background: Neurological illness can produce a disorganization of behavior, including verbal disinhibition, despite apparent preserved intelligence. Neuropsychological tests of such behavioral control mechanisms may predict real-world performance of healthy people, such as success or misbehavior in educational contexts.
Method: In two separate studies, we examined how the Hayling Test of verbal response suppression predicts grades and classroom misbehavior.
Results: Verbal suppression errors and spontaneous strategy use were significant predictors of undergraduate grades. Using a modified version of the Hayling Test designed to reduce strategic responding with high school students (mean age 16), higher grades were predicted by shorter response suppression latencies and better working memory scores, and classroom misbehavior was predicted by lower working memory scores.
Conclusion: Verbal response suppression and spontaneous strategy use, both closely linked to disorganized behavior in neuropsychological patients, predict academic achievement but seem unrelated to classroom misbehavior, which is associated with weakness in working memory.

1. Introduction

Students vary on many dimensions, including their motivation, personality and of particular relevance here, their cognitive abilities. In fact, based on a review of meta-analyses it has been estimated that variation in academic achievement at university level is around 50% dependent on individual differences between the students, with only about 20–25% dependent on differences between professors, such as attitude or teaching styles [1].

Traditionally, psychologists have attempted to understand this between-student variation through the concept of general cognitive ability, i.e. intelligence. However, although intelligence is a good predictor of academic ability in younger learners [2,3], it is rather limited in understanding achievement at higher levels such as late high school and university study. In fact, some studies have failed to detect any correlation between intelligence test scores and grades achieved by high school [4] or undergraduate students [5,6].

One alternative approach has been from cognitive psychology, in which attempts are made to link specific cognitive processes to academic performance, usually measured through grades. This research has particularly linked subprocesses of working memory to success in certain subjects. For example, visuospatial working memory ability appears to be an important individual difference that predicts how efficiently young people learn mathematical skills and concepts, but phonological loop processes are more important in later use of the mathematical skills [7]. Phonological processes are also closely linked to language grades [8]. This corpus of research, mainly conducted in young children, has suggested that specific features of working memory have domain-specific contributions to educational performance. Nevertheless, working memory is very closely associated with general intelligence, with about two-thirds of the variance in children's working memory ability attributable to general intelligence [9]. Furthermore, there is evidence that working memory is a poor predictor of academic performance of older learners, such as undergraduate students [10].

An alternative approach comes from neuropsychology. The frontal lobes have long been associated with adaptive goal-directed behavior and impaired real-life functioning in disorders that compromise frontal lobe functioning [11,12]. Interestingly, there have been multiple

* Corresponding author.
E-mail addresses: g.c.pluck@gmail.com (G. Pluck), mikarolysc@usfq.edu.ec (M.I. Karolys), mmontano@estud.usfq.edu.ec (M.E. Montaño-Córdova), p.meza.17@ucl.ac.uk (P. Almeida-Meza).

https://doi.org/10.1016/j.tine.2019.100121
Received 16 April 2019; Received in revised form 28 July 2019; Accepted 9 October 2019
Available online 10 October 2019
2211-9493/

reports of adult patients with damage to the frontal lobes who became grossly disorganized in their lives. This disorganization resulted in frequent debts, dismissal from employment, dropping out of higher education, etc. Nevertheless, despite their real-life failures, the patients maintained high intelligence as shown by normal or in several cases, well above normal, IQ scores [13–17]. This suggests that some neuropsychological abilities associated with frontal lobe functioning may be useful in measuring abilities that are relatively distinct from intelligence yet still contribute to real-life successful adaptive behavior.

The cognitive and neuropsychological approaches are not mutually exclusive, as there is significant cross-fertilization between the perspectives, particularly regarding working memory and executive functions in general. The latter covers a wide range of abilities but is often thought to be summarized by three basic factors: working memory, shifting and inhibition [18]. One feature of executive function, verbal response suppression, may be of particular relevance in understanding academic attainment in older learners (e.g. adolescent students and undergraduates). This can be seen as a form of inhibition, but unlike most cognitive assessments of inhibition (e.g. Go/No-go, Stop-signal), it involves the suppression of verbal responses. Furthermore, neuropsychological evidence suggests that the biological basis of the form of verbal response suppression discussed here is different to that of other forms of inhibition, such as that measured by the Stroop task [19]. However, in contrast to the numerous studies of working memory as a predictor of academic achievement, there are very few studies of verbal response suppression as a predictor.

The principal neuropsychological test for verbal response suppression is the Hayling Test [20,21]. This was developed to measure the disorganization of behavior of patients with damage to the frontal lobes, one manifestation of which is verbal disinhibition. This is characterized by, for example, lewdness, jocularity, and insensitivity to the social context [22]. Indeed, verbal disinhibition can be caused by a variety of illnesses that damage the frontal cortex [23–27]. Basically, a single trial on the Hayling Test involves participants completing a sentence that lacks the final word by saying a word that makes no sense, i.e. suppressing the prepotent response. Neuropsychological studies have consistently linked performance of this test to the integrity of the frontal lobes, in particular, the right prefrontal cortex [20,28–30]. It is also notable that impaired Hayling Test performance in frontal patients is independent of changes to general intelligence [19,29], and in healthy adults, Hayling Test performance is not correlated with IQ [10]. This further supports the proposition that assessments such as the Hayling Test may provide an alternative neurocognitive perspective to understanding academic achievement.

Indeed, we previously studied the Hayling Test and several other neuropsychological tests linked to frontal lobe functioning, as well as a measure of intelligence, as potential predictors of Grade Point Average scores (GPA) in university students. This revealed that the Hayling Test outperformed all of the other neuropsychological tests, including one of working memory, in predicting GPA [10]. Furthermore, Hayling Test performance was better than intelligence test scores and explained a significant amount of additional variance in GPA to that explained by the intelligence test. This hints at an alternative approach to understanding academic achievement of older learners such as undergraduates. Rather than intelligence or working memory, mechanisms linked to behavioral control such as response suppression may be more important in complex environments such as late high school or higher education. This is because these educational contexts require significant self-directedness, initiative, etc. In support of this approach, a study that used the Behavior Rating Inventory of Executive Function-Adult Version [31] with undergraduate students reported that the Behavioral Regulation Index was consistently negatively correlated with different indexes of academic attainment [32]. In fact, self-discipline, a trait closely linked to frontal-lobe functioning [33], explains more than twice the variation in academic performance of adolescents than intelligence [34].

The Hayling Test may simply be a good measure of this frontally-mediated behavior regulation. This makes sense if the Hayling Test is really measuring the ability to suppress contextually inappropriate responses. However, it has been suggested that the reason that Hayling Test performance is linked to frontal lobe functioning is that performance is open to strategy use, which is itself a prefrontal-linked function [35]. While healthy individuals tend to spontaneously apply strategies to perform the Hayling Test, frontal lobe damaged patients do this less often and less efficiently [28]. This suggests that the Hayling Test may actually be measuring spontaneous strategy use, and the previously observed association with GPA may be a reflection of that, not of response suppression *per se*.

In this research, we attempted to explore in more detail the relationship between Hayling Test performance and academic achievement in older adolescents and young adults. Two separate studies are reported. In the first, we attempted to replicate our previous finding that Hayling Test performance is a good predictor of GPA in university students. In the second study, we examined various other issues related to Hayling Test performance and academic achievement. These include the contributions of response suppression and spontaneous strategy use, whether the effect generalizes from undergraduate students to high school students, and the role of general verbal ability. Furthermore, as it is potentially measuring behavioral regulation, we investigated whether Hayling Test performance is related to classroom misbehavior.

2. Study 1

The principal aim of this study was to replicate our previous finding that verbal response suppression is associated with academic achievement in university students [10]. A secondary aim was to assess the two other principal executive functions, working memory and switching. As the Hayling test is a verbal test, we included working memory and switching tests that also use verbal material. Furthermore, as Robinson et al. [28] have shown that the poor response suppression as measured by error scores or response times may actually be caused by a failure to spontaneously use strategies, at least in frontal lobe patients, we also coded all responses for strategy use using the same criteria as Robinson et al. (2015).

2.1. Method

2.1.1. Participants

A sample of 50 undergraduate students (24 female, 48.00%) at Universidad San Francisco de Quito, Ecuador, was recruited. Due to an error, age was not recorded, but all were in their first semester of study, and the modal age would be highly likely to be 18 as that is the typical age of starting university in Ecuador. The most common majors being studied were psychology ($n = 17$, 34.00%), followed by industrial engineering ($n = 9$, 18.00%) and mechanical engineering ($n = 6$, 12.00%).

2.1.2. Cognitive assessments

All participants were assessed with three cognitive tests. The Hayling Test is comprised of two parts, 1 and 2. In both parts, 15 sentences are read to the participant by the experimenter. However, in each sentence, the final word is missing, and it is the task of the participants to quickly say a word that completes it [20,21]. The sentence context constrains responses, such as “The captain should be the last to leave a sinking…”. Part 1 is simply an initiation task with no suppression component in which participants complete sentences as quickly as possible with appropriate words. However, as we were only interested in suppression performance, in this study, we used only the Hayling Test Part 2. In this section, the participants are required to quickly complete each sentence with a word that makes no sense, such as ‘banana’. Therefore, they have to suppress any prepotent responses primed by the sentence. A Spanish-language version of the Hayling Test was used here, which has previously been used for research [10,36].

Response times (in seconds) on the Hayling Test are recorded with a stopwatch from the last word spoken by the experimenter to the onset of the response by the participant. Regarding the error score for Part 2 (as used here), if the participant performs well, producing a word completely unrelated to the sentence, they receive zero error points. However, if they produce a word that sensibly completes the sentence, which is what they were instructed not to do, they receive three error points. An intermediate response that does not complete the sentence reasonably, but still shows an influence of semantic context receives one error point. Such one-point responses are words that are semantically related to, or opposite to, the prepotent response. The total of these scores across the 15 trials gives the suppression error score. High scores indicate poor performance (maximum 45 points).

In addition to the main performance measures, we coded all responses in terms of the response type classifications of Robinson et al. [28]. These can be used to detect specific strategies used to produce correct (nonsense) responses. In this analysis, we are only interested in the strategies used to produce correct (0 error point) responses. These are classified as being 'correct and visible', 'correct and semantic to a previous response', and 'correct and both visible and semantic to a previous response'. We used this to tally counts of visible strategic responses, semantic strategic responses, and total strategic responses (visual, semantic, or both). As there were 15 trials, the maximum possible frequency for each strategy type is 15 (i.e. the participant used that particular strategy on every trial).

As a measure of verbal working memory, we used the Counting Span Test [37]. This is similar to a simple digit span assessment but contains elements that provide interference, particularly tasking working memory. Participants are asked to count aloud dark blue circles among arrays of light blue circles and dark blue squares (i.e. perform a conjunctive search). At the end of the count, they then say aloud the total before the next trial begins. After several trials, they are asked to recall the list of total counts in the correct order. The length of the sequences (i.e. number of totals to recall) begins at two and increases gradually to five, with two trials at each length. For scoring, we used unit-weighted partial credit scoring, as this has been shown to be a good measure of performance in working memory span tasks [37].

As a measure of switching ability, we used the Trail Making Test provided in the Delis-Kaplan Executive Function System [38]. As the baseline measure of speed of responding, we used the average time to complete the Number Sequencing and Letter Sequencing conditions. The variable of interest is calculated by dividing the time taken on the Switching condition by the baseline speed of responding time. This has been validated as the best measure of switching ability on the Trail Making Test [39].

2.1.3. Procedure

All participants provided written informed consent, in accordance with the institutional ethics committee approved protocol. Participants were tested individually on all the tests listed above, in the order listed. Assessment of each participant took approximately 15 minutes. After the semester in which the data was collected, the University computer systems were accessed to obtain the average GPA for the semester. GPA in this university ranges from 0–4 (best).

2.1.4. Statistical analysis

Data distributions were checked with Shapiro-Wilk tests. Any data distributions that were significantly nonnormal were transformed to improve their distributions with the RANKIT method [40]. Pearson correlations were used for all bivariate analyses and were two-tailed with a significance level of 0.05.

2.2. Results and discussion of Study 1

The correlation coefficients between cognitive variables and GPA, as well as mean scores, are shown in Table 1. There were no significant correlations between GPA and working memory capacity as measured by the Counting Span Task. Nor was there an association with switching ability as measured by the Trail Making Test. However, there were significant correlations between Hayling Test Part 2 performance and GPA. There was a significant negative correlation with the suppression error score, indicating that students with lower GPA tended to make more verbal suppression errors. In this study, we did not include the Hayling Test Part 1, which is mainly used to measure initiation time as the task requirement is to produce sensible words to complete the sentences. This omission was because we were interested only in suppression. However, it is likely that Part 1 produces a response-set of predictable word sentence completion, which may increase the error rate on Part 2. Our omission of Part 1 may, therefore, have reduced the frequency of suppression errors, and should be considered a limitation of the current study.

Table 1
Mean (+SD) scores on the cognitive tests by the university sample and their correlations with grades.

	Scale /maximum possible score	Mean (+SD)	*r*
Counting Span	Total score /8	5.859 (1.090)	.100
Trail Making Test	Switching cost	2.028 (0.790)	.009
Hayling Test	Median RT	1.106 (0.646)	−.170
	Suppression total error score /45	4.200 (3.928)	−.301*
	Correct with a semantic strategy /15	0.820 (0.941)	−.057
	Correct with a visual strategy /15	5.380 (3.697)	.224
	Correct with any strategy /15	6.580 (3.742)	.294*

* $p < .05$.

In addition to the correlation between GPA and suppression error scores, there was also a significant positive correlation for the count of responses that used a strategy. Examining the table, this seems to be mainly driven by the association between GPA and visual strategies (such as naming objects visible during testing) rather than semantic strategies (such as repeating words from previous sentences). To confirm this, we performed a separate analysis on the total number of strategic responses calculated as those with a clear visible strategy plus those which were ambiguous (both verbal and visual). This new total correlated with GPA at $r = .274$, $p = .054$. This showed a trend for statistical significance, and the size of the correlation is clearly greater than that between GPA and the total number of semantic strategy responses.

These correlations suggest that the actual ability measured by the Hayling Test Part 2 that correlates with GPA may be spontaneous strategy application and not response suppression *per se*. On the one hand, the current result does replicate the previous finding that Hayling Test Part 2 performance is a significant correlate of academic performance in university students, and it appears to have a stronger relationship than other cognitive tests [10]. On the other hand, the pattern of correlations involving strategic responses suggests that, although the Hayling Test Part 2 is generally used as a measure of response suppression, some participants spontaneously use strategies to achieve low error scores. It may actually be this strategy application which is associated with academic achievement. Nevertheless, the observation that spontaneous strategy application during testing is a significant correlate of GPA adds to our understanding of the role of executive functions in academic attainment.

The issue of response suppression versus spontaneous strategy use is a consequence of the way that the Hayling Test is administered. As alluded to in the introduction, there are several features of the Hayling Test that mean participants can approach it in different ways, or use different cognitive abilities, to achieve the required performance outcome. One of these features is that in the standard administration there are 15 trials of initiation (Part 1) in which the task is to complete the sentences normally, followed by 15 trials of suppression (Part 2) in which the task is to produce a nonsense response. This allows participants to potentially prepare strategies, as they know the task demands well in advance. A second feature is that participants can potentially pick a nonsense word and hold it in working memory, ready to say at the appropriate moment. From the perspective of the participants, picking a word such as 'mushroom' will probably allow them to produce a nonsense response to whatever sentence the experimenter reads to them. In Study 2, we used a standard Hayling Test, plus a modified Hayling Test designed to block strategic responding such as holding words in memory ready for use. This allows us to better understand the contributions of strategy use and response suppression to academic attainment. With the modified Hayling Test, if the correlation with GPA is maintained despite the changes to the task to block strategy use, there probably really is a relationship between response suppression and academic achievement.

A further issue addressed in Study 2 is whether this relationship between academic achievement and Hayling Test performance is specific to adults in higher education. In several studies with school children, working memory performance has been found to be the best predictor of academic grades, particularly in language and mathematics [2,7,8,41,42]. However, in this study and in our previous research with university students [10], working memory ability has been found to have no association with grades, while supposed response suppression (Hayling Test Part 2 errors) does. In study 2, we used a high school sample of participants (age range 15–17). As these participants performed both a standard Hayling Test and the novel modified Hayling Test, in addition to testing the differential influence of response suppression/strategy application, we were able to explore the generalizability of the basic results to school performance.

That response suppression may be linked to academic achievement at university level, whereas other executive functions such as switching and working memory are not, as observed in the current study, requires some explanation. One factor may be that the mechanism through which working memory contributes to academic achievement in younger learners is that better working memory directly facilitates comprehension and calculation, i.e. students with better working memory ability can hold and process information more efficiently in real-time [7]. However, there is evidence that response inhibition may be associated with school grades through longer-term mechanisms, such as good inhibition ability being a feature of greater self-control and self-directed learning [43]. Further support for this comes from the observation in children that while working memory ability may be specifically associated with reading and mathematical ability, 'executive functions' (i.e. inhibition, switching and planning) may have a non-specific association with academic achievement in general [41]. This has intuitive attraction as a mechanism for why young people with good response suppression/inhibition may also tend to achieve better grades. The emphasis on response suppression as self-control, also suggests that associations between suppression task performance may go beyond associations with grades and may also be linked to classroom behavior. In Study 2, in addition to investigating relationships between Hayling Test performance and grades, we also examine whether it is associated with classroom misbehavior in school-age adolescents.

Finally, as the Hayling Test is ostensibly a test of verbal response suppression, and performance has been shown to correlate substantially with verbal ability in young adults [44], in Study 2 we also investigated whether general verbal ability could explain why students with higher GPA tend to have better performance on verbal response suppression.

3. Study 2

3.1. Method

3.1.1. Participants

One-hundred and six high-school students were recruited over two semesters at a state-run school in Quito, Ecuador. Mean age in years, calculated from date of birth and interview date, was 16.077 (SD = 0.551, range 14.987–16.994). Fifty-five (55.887 %) were male, and the majority (103/106, 97.170%) identified as being ethnically 'mestizo' (mixed European and Native American ancestry). A preliminary analysis of the first 66 participants has previously been published as a Spanish-language conference proceeding [45].

3.1.2. Materials

3.1.2.1. Assessment of misbehavior. To measure classroom behavior, we developed a brief scale containing five items, known hereafter as the Misbehavior Scale. The five items included in the Misbehavior Scale pertain to issues identified as particular concerns by one of us (DV), who was a teacher at the school where the data was collected. The items asked about academic dishonesty, interruptions, lateness to class, aggression to fellow students and aggression to teachers. A custom scale was developed because student behavior scales available in Spanish and validated for this country were not available. Furthermore, scales of classroom behavior that are available in English and have been previously associated with cognitive performance, such as Conner's 3 Teacher's Short Form [46], focus on the impressions of teachers concerning impulsive behavior, executive functioning, etc. of their students. The associations with performance on executive function tests are therefore not surprising. In this research, we wished to examine associations between Hayling Test performance and actual reported classroom misbehavior.

The questions asked (in Spanish) and the English translations are shown in Table 2. Teachers were asked to rate, on a Likert scale from 1–4 (never to always), how each student tends to behave in their particular class on those 5 behaviors.

3.1.2.2. Assessment of cognitive abilities. To measure verbal response suppression, we used the Hayling Test Parts 1 and 2. The original test was published in English, but we used a published Spanish version that has been shown to have good reliability [47].

We will refer to this in the text that follows as the standard Hayling Test, to distinguish it from a modified version that we also used, described later. The standard Hayling Test has two parts. On Hayling Part 1, the task is simply to complete sentences with an appropriate word. The test measures initiation, but there is no suppression component. Due to the simplicity of this task, errors are very rare; the variable of interest is the response time as this gives an indication of initiation speed.

After Hayling Part 1 is completed, participants perform the Hayling

Table 2
Questions in the Misbehavior Scale.

Item	Original Spanish	English translation
1	Realiza actos de deshonestidad académica	Performs acts of academic dishonesty
2	Interrumpe la clase (usa su celular, se levanta, hace ruido, etc.)	Interrupts the class (uses their cell phone, gets up, makes noise, etc.)
3	Llega tarde a clase	Arrives late to class
4	Agrede física o verbalmente a sus compañeros/as (ejemplo: golpes, insultos sexistas o racistas)	Assaults classmates physically or verbally (examples: punches, sexist or racist slurs)
5	Es irrespetuoso /a o agresivo/a con su profesor/a	Is disrespectful or aggressive to the teacher

Part 2. The administration of this and stimulus materials are equivalent to those in Part 1, i.e. 15 sentences in which the last word is missing. However, as described previously in Study 1, the task is now to suppress the prepotent obvious response and say something that makes no sense in the context of the sentence. We calculated the suppression error score and divided that by the number of trials to obtain an error score per trial measure. Although the Hayling Test is usually used with adult populations, its use has recently been extended to adolescents [48].

As the Hayling Test is basically a measure of verbal response initiation (Part 1) and verbal response suppression (Part 2), we also included a measure of general verbal ability. This was the Vocabulary subscale from the Wechsler Intelligence Scale for Children-IV (WISC) Spanish Version [49]. This involves participants providing definitions for 36 words of graded difficulty. This is a widely validated and reliable test of linguistic ability for use with children and adolescents.

We also included a measure of verbal working memory capacity: The Reading Span Test. Good performance on this test is associated with shorter inhibition response times on the Hayling Test [44], suggesting performance on the latter may be partly driven by verbal working memory ability, perhaps through facilitating the strategic storage of verbal responses in working memory (discussed further below). The Reading Span Test involves participants reading sentences aloud. In addition, they are required to remember the last word of each sentence until a point when they are asked to recall the last words. There are several sets of sentences so that the number of words is gradually increased, from two sentences (two words to recall) up to five sentences (five words to recall). This is considered to challenge executive control as participants need to maintain the to-be-recalled words while simultaneously reading aloud, so it is described as a working memory span task. Performance on the Reading Span Test is highly correlated with various measures of linguistic comprehension [42]. We used a Spanish language version [50], which is presented on a computer screen. As with the Counting Span Task used in Study 1, we used unit-weighted partial credit scoring [37].

Finally, we also included a new experimental version of the Hayling Test, referred to hereafter as the modified Hayling Test (see Fig. 1). The test is essentially the same as the standard test. Participants must complete sentences with either an appropriate word or an inappropriate word (as they do in the standard Hayling Test Parts 1 and 2 respectively). The aim of this modified version is to reduce preplanning and storage of responses ready for utterance at the start of the response time recording. There are three key modifications to that end. The first is that whether an initiation or a suppression response is required is not told to the participants until the start of the response time recording. The standard Hayling Test has 15 initiation trials followed by 15 suppression trials. In the modified Hayling Test, there are 9 initiation trials and 14 suppression trials presented in a fixed pseudo-random sequence. The sentences to be completed are presented on a computer screen in a PowerPoint file. At the start of the response time recording, the experimenter advances the slide which then shows either a green tick mark accompanied by a chime sound or a red cross accompanied by a beep. The participant is instructed to complete sentences with an appropriate word when presented with a tick mark and chime or an inappropriate word when presented with a cross and beep. This modification, therefore, removes the predictability of which trial is coming next, which is inherent in the standard Hayling Test.

The other modifications are designed to increase working memory load so that it becomes difficult to maintain possible responses for use in coming trials. This was achieved by adding a working memory span task component. Mixed within the 9 initiation trials and the 14 suppression trials are 7 trials in which the sentence does not have to be completed by the participant. On those 7 trials, at the point at which an initiation or suppression response is usually prompted (with a tick or a cross), the word completing the sentence is presented to the participants on the computer screen as a written word (in purple type). The task of the participant is to hold this word in memory until prompted to recall later in the test. There are either three or four intervening trials between word presentation and word recall, in an unpredictable sequence. Effectively this means that throughout the task, participants are attempting to maintain a single word in short-term memory and that word changes every few trials.

The final modification of the Hayling Test is that the participant must read aloud the entire sentence. In the standard Hayling Test, the experimenter reads the incomplete sentences aloud, and the participant gives only the final words to complete them. Making the participants read the sentences plus responses makes it difficult for them to store responses for use in coming trials. The modified Hayling Test can be seen as a combination of the standard Hayling Test Parts 1 and 2 (measuring initiation and suppression) with a Reading Span Task (measuring verbal working memory span).

The modified Hayling Test presented here contains 14 suppression trials, 9 initiation trails and 7 memory trials. Therefore, it has the same number of trials overall as the standard Hayling Test. The 30 sentences used here in the modified version are different from the 30 sentences used in the standard Hayling Test described above (developed in Spain) and were taken from a version developed for use in Cuba [36].

3.1.3. Procedure

All 106 participants were recruited from a single school. The data was collected in two consecutive academic years to provide a sufficient sample size. The participants were all students at the school and provided written assent to participate, and a parent or guardian additionally provided written consent for their participation, in accordance with the institutional ethics committee approved protocol. Each participant was assessed individually in a quiet office at the school. Basic demographic data were collected, followed by the various cognitive tests. Participants were first tested on the WISC Vocabulary Test. Following this they were tested on Parts 1, then 2, of the standard Hayling Test. This was administered in its standard format with response times recorded on a stopwatch. Next, all participants completed the Reading Span Test, which was presented on a laptop computer screen. Finally, the modified Hayling Test was administered. This was given last as it contains elements of the previously administered tests (standard Hayling Test and Reading Span Test) and so participants were already familiar with the task demands. The test was administered as a PowerPoint file on a laptop computer. At the appropriate time, when a participant was to complete a sentence, the experimenter would advance the slide that indicates the type of trial (memorize, initiation, or suppression). Each application of the modified Hayling Test was audio recorded. The experimenter listened to the recording after the participant had left the test session and transcribed the content of the responses (utterances) and response times on a standard record form. The participants were then debriefed, thanked for their participation, and given a small gift of a highlighter pen worth about $1.50.

At the end of the semester in which the cognitive data was collected, we took the final class grades for each participant for that semester from the school database. Additionally, the verbal responses on the standard and modified Hayling tests were scored for strategy use.

During the semester of data collection, we asked the teachers in four classes to complete the Misbehavior Scale for each participant. Therefore, we had four Misbehavior Scales completed for each participant. This allowed us to take an average of misbehavior scores to minimize variation caused by situational factors (i.e. students behaving better in some classes than others).

3.1.4. Statistical analysis

Data were processed with SPSS Version 23. For time-based measures, median response times were calculated for each participant in each condition. For the standard Hayling Test and modified Hayling Test, the two versions have different numbers of suppression trials (15 in the standard and 14 in the modified version). In order to compare performance between these two versions, absolute scores for errors and

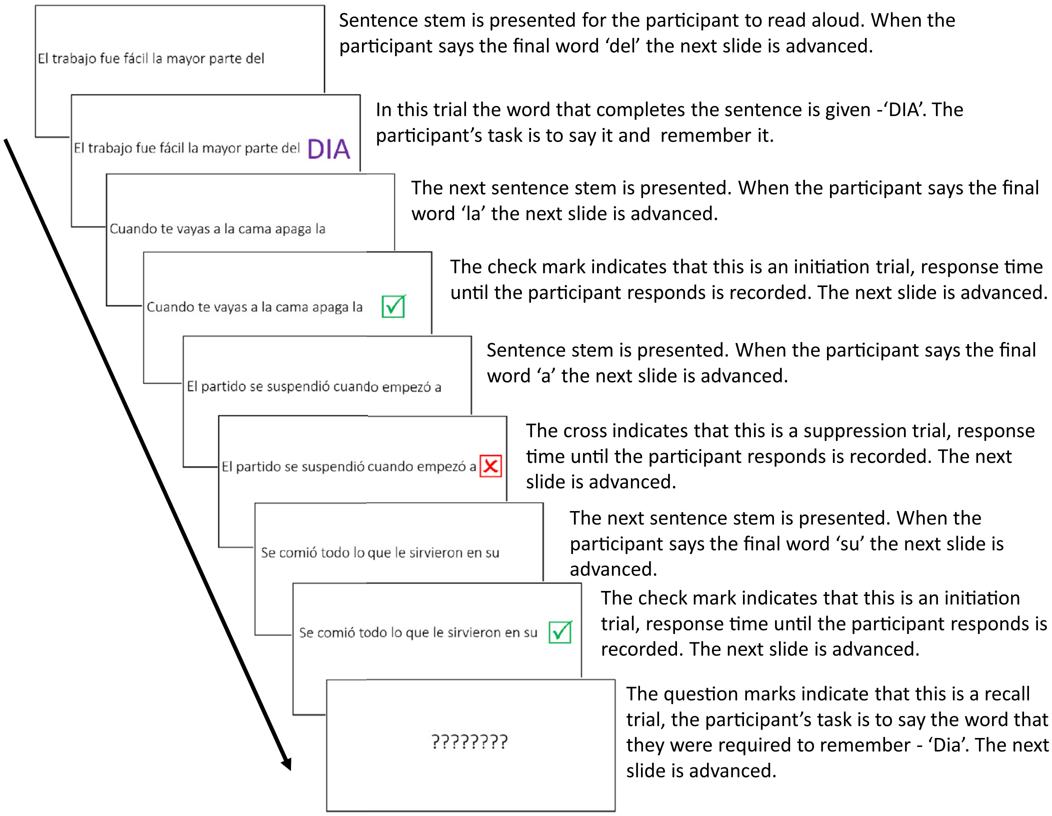

Fig. 1. Schematic representation of the sequence of stimuli and trials in the modified Hayling Test. The sample stimuli show how 4 trials were administered (1 working memory, 1 suppression, and 2 initiation).

strategy counts were not used. Instead, scores per trial were calculated for each participant. Consequently, the maximum total per item score for strategy counts are all one point, and the maximum total error scores per item are three points. As examples, if a participant made five correct responses on the modified Hayling Test which exhibited visual strategies, their adjusted score would be 5/14 = 0.357; if a participant achieved a total error score of 20 points on the standard Hayling Test, their adjusted error score would be 20/15 = 1.333 (or 1.333 error points per trial).

For the other cognitive tests, the raw total scores were entered into SPSS. Age-adjusted scores were only available for one of the tests (WISC Vocabulary) but are not based on the Ecuadorian population, and so were not used. Instead, age is entered as a covariate in all correlation and regression comparisons. To prevent the effects of extreme scores excessively influencing regression analyses, scores were Winsorized at 3 SDs. As in Study 1, data distributions were checked with Shapiro-Wilk tests, and any data distributions that were significantly nonnormal were transformed with the RANKIT method [40]. These transformed scores were used for the correlational analyses (Pearson coefficients). However, for regression analyses, untransformed scores were used, but the distributions of the residuals were checked for normality of distributions with Shapiro-Wilk tests. For within-subject analyses, when the distributions of data remained nonnormal even after transformation, non-parametric analyses were performed with Wilcoxon matched-pairs sign-rank tests (for which the standardized test statistic is reported). Otherwise, for normally distributed data, ANOVA was employed. All analyses were two-tailed with a significance level of 0.05.

3.2. Results and discussion of Study 2

In the first analysis, we examined the statistical properties of the scale of behavior used in the study - The Misbehavior Scale. The first three items, Dishonesty, Interruptions, and Punctuality, had approximately normal distributions based on skewness and kurtosis. However, the final two items, Aggression to Students and Aggression to Teachers both demonstrated positively skewed and leptokurtic distributions. For this reason, internal consistency was estimated by calculating the mean inter-item Spearman correlation. This was found to be 0.501 (range 0.278 to 0.662), a mean and range suggesting good internal consistency and unidimensionality [51].

Next, we examined performance on the modified Hayling Test. The first step was to compare performance of the modified Hayling Test with the standard Hayling Test, in order to see how the modifications changed performance. The variables of interest are shown in Table 3 and are all calculated and reported as per trial scores, to adjust for the different number of trials in the modified and standard Hayling Tests. For example, the reported figure of 0.035 for semantic strategies on the standard Hayling Test signifies that the mean average for the group was to make 0.035 responses that appeared to use a semantic strategy for each sentence completed, or in other words a mean average 0.525 semantic strategies over the whole 15 sentences (0.035 × 15). Compared to the standard Hayling Test, the modified Hayling Test was associated with longer initiation times (Wilcoxon Signed-Rank Test = 7.282, $p < .001$), longer suppression times ($F(1,104) = 114.809$, $p < .001$, $d = 1.041$), longer adjusted suppression times (suppression times minus initiation times), $F(1,104) = 20.591$, $p < .001$, $d = 0.441$), and higher suppression error scores (Wilcoxon Signed-Rank Test = 8.126, $p < .001$). These increases suggest that the modified Hayling test degraded performance for both initiation and suppression, in comparison to the standard Hayling Test, probably because of the increased cognitive load.

Of particular interest here is whether spontaneous strategy application was reduced as that was the aim of the modifications. This was confirmed. Overall, participants made fewer responses that appeared to use a strategy on the modified Hayling Test, in comparison with their performance in the standard version (Wilcoxon Signed-Rank Test = -

Table 3
Mean (+SD) scores on the cognitive tests by the high school sample and their correlations with grades and misbehavior.

	Scale /maximum possible score	Mean (SD)	Correlation with Grades	Misbehavior
Vocabulary	Total score /68	38.566 (6.749)	.000	.151
Reading Span	Total score /15	8.602 (1.408)	.184	−.063
standard Hayling Test	Initiation time	0.766 (0.134)	−.117	−.037
	Suppression time	1.426 (0.692)	−.066	−.020
	Suppression time – initiation time	0.658 (0.650)	−.036	−.033
	Suppression error score (per trial) / 3	0.383 (0.444)	−.081	.011
	Correct with a semantic strategy (per trial) /1	0.035 (0.067)	−.200*	.112
	Correct with a visual strategy (per trial) /1	0.305 (0.258)	.071	.029
	Correct with any strategy (per trial) /1	0.345 (0.258)	−.022	.071
modified Hayling Test	Initiation time	1.088 (0.380)***	.085	−.176
	Suppression time	2.172 (0.723)***	−.222*	−.068
	Suppression time – initiation time	1.082 (0.690)***	−.269**	−.016
	Suppression error score (per trial) /3	1.100 (0.517)***	.001	−.036
	Correct with a semantic strategy (per trial) /1	0.030 (0.056)	−.061	.068
	Correct with a visual strategy (per trial) /1	0.138 (0.156)***	.078	.037
	Correct with any strategy (per trial) /1	0.170 (0.167)***	.063	.030
	Memory span /7	4.745 (0.986)	.164	−.228*

*$p < .05$, **$p < .01$, $p < .001$. In the Mean (SD) column, the *s indicate significant differences between performance on the standard and modified Hayling Tests. The Strategy scores are the tally of trials in which the strategy was observed divided by the number of trials (15 in the standard Hayling Test and 14 in the modified Hayling Test). The Suppression error scores are similarly adjusted to be per trial.

6.866, $p < .001$).

However, although the modifications significantly reduced the number of visual strategies detected (Wilcoxon Signed-Rank Test = -6.805, $p < .001$), there was no significant difference between versions of the Hayling test for semantic strategies (Wilcoxon Signed-Rank Test = 0.470, $p = .638$). It appears that although the modified Hayling Test used here reduced strategy use overall, it did so by reducing the frequency of visual strategy use. An explanation for this is tentatively suggested. The extra cognitive load probably reduced visual strategy use, as it requires attention and effort. For example, to use a visual strategy, one has to look around the room and pick something, such as the window, and then use the word ‘window’ to complete the sentence. Semantic strategy use is rather different, and would involve, for example, saying the word ‘apple’ to complete a sentence, when a previous response had already been semantically related to that, such as ‘orange’. Perhaps, such semantically related responses occur not through a deliberate search for a response, but through semantic priming. This phenomenon occurs when exposure to a word makes that or related words more likely to be produced or recognized, and generally occurs automatically, without effortful control [52].

In the next analyses, we examined whether classroom misbehavior or academic achievement are correlated with performance on the standard cognitive tests (WISC Vocabulary, Reading Span, standard Hayling Test). These are also shown in Table 3. For academic grades, there was only one significant correlation. This was for the number of correct responses on the Hayling Part 2 in which semantic strategies were used. Interestingly, this was a negative correlation, suggesting school students with higher grades were less likely to use that strategy than students with lower grades. One interpretation of this may be that semantic responses on the Hayling Part 2 may actually indicate relatively poor error monitoring, i.e. difficulty preventing oneself from producing automatic semantic primed responses. This is similar to the explanation proposed for why patients with damage to the right prefrontal cortex make more subtle semantically related responses on the Hayling Test than patients with left prefrontal damage [28], as the right prefrontal cortex has been previously identified to have a role in error monitoring [53].

This negative correlation between semantically related strategies in the standard Hayling Test and academic grades is opposite to the correlation found in Study 1, in which there was a positive correlation between any strategy use and grades of university students, which seemed mainly driven by the use of visual strategies. It may be that the more cognitively-able students are more likely to use effortful strategies such as picking a visible object when performing the standard Hayling Test, or they are better able to monitor and prevent semantically primed responses. These cognitively-able students may also tend to have higher GPA.

It is also notable that there was no significant association between grades and the Hayling Part 2 suppression error score. Therefore, we failed to replicate the finding from Study 1. Regarding the Misbehavior Scale, there were no significant correlations with any of the standard cognitive measures (i.e. Vocabulary, Reading Span, standard Hayling Test). Overall, the associations between grades and classroom misbehavior and the standard cognitive tests revealed only one significant association, between overall grades and semantic strategy use on the Hayling Test Part 2.

We turn now to an examination of the correlations of educational measures with the modified Hayling Test, which was developed specifically for this research. This combined task essentially has four different concurrent measures; working memory score, initiation response time, suppression response time, and suppression error score. In addition, as is usual with the Hayling Test, we calculated the suppression time minus initiation time score. The correlations between these scores and the target variables of grades and classroom misbehavior are also shown in Table 3. Unlike in Study 1, there are no significant associations between educational variables and strategy use, perhaps due to the measures taken to reduce such strategies in the modified version of the test. Nor is there an association with the working memory span score. In contrast, there are now negative correlations between grades and response suppression times, indicating that students with higher grades had more efficient verbal response suppression. This is observable in the two different suppression trial response times, the simple measure of absolute time to respond, and the version that is adjusted by subtracting the median response time on initiation trials.

On the Misbehavior Scale, as with the standard Hayling Test, the modified Hayling Test produced no associations with verbal response suppression. Interestingly though, performance on the embedded working memory span task was negatively correlated with classroom misbehavior. This indicates that poorer working memory performance is associated with more classroom misbehavior. That this correlation was only revealed in the modified Hayling Test, and not with the Reading Span Test, may be due to the extra cognitive load induced in the former. It is worth remembering that the modified Hayling Test is considerably more challenging to perform than either the standard

Hayling Test or the Reading Span Test, as it essentially requires participants to perform aspects of both simultaneously.

Overall, the results of the correlational analyses with the modified Hayling Test provide evidence that grades of the high school students are correlated with their verbal response suppression ability. On the other hand, misbehavior is correlated with their verbal working memory performance.

However, all of the modified Hayling measures are likely to interact on suppression performance as they are derived from the same test. For this reason, linear regression analysis was also used to reassess the results of the correlational analyses. The benefit of linear regression analysis is that the association of each independent variable (cognitive performance) with the dependent variable (grades or misbehavior) can be assessed, with the other independent variables held constant (entered simultaneously). This essentially checks whether the results observed in the simple correlational analyses are due to, or masked by, associations between the different independent variables, such as speed-accuracy trade-offs.

For the linear regression analysis predicting grades, we included initiation times, suppression times, suppression error scores, and working memory span performance as independent variables of particular interest. We also included age and WAIS Vocabulary scores to control for general verbal ability. The results are shown in Table 4.

With all variables entered, the regression produced a significant model predicting grades, $F(6, 99) = 2.932$, $p = .011$, with an R of 0.388 (adjusted $R^2 = .099$). However, within the model, only suppression response times and working memory scores were significant predictors of overall grades. Higher grades were associated with lower response suppression latencies and higher working memory scores, although initiation times were approaching significance at $p = .055$. This regression result confirms the results with bivariate correlations. However, working memory additionally emerged as a predictor of grades, whereas the correlation was not significant. As the regression method included WAIS Vocabulary scores, variation in language skills is unlikely to be the cause of the associations between Hayling suppression times and working memory scores.

Furthermore, this regression model appears to have acceptable statistical properties. The Variance Inflation Factors (VIFs) were all low (range 1.025–1.223) indicating multicollinearity is not a problem. In addition, the distribution of residuals from the model did not differ significantly from normality (Shapiro-Wilk test = 0.983, $p = .209$).

For the linear regression predicting classroom misbehavior, the same procedure was followed with the same independent variables. The results are shown in Table 5. This again produced a significant model predicting classroom misbehavior, $F(6,99) = 2.313$, $p = .039$, with an R of 0.351 (adjusted $R^2 = 0.070$). However, within this model, response suppression was not a significant predictor of classroom misbehavior. Age appeared as a significant predictor; younger age associated with more misbehavior. Working memory scores remained significantly associated with the dependent variable, indicating lower working memory performance was associated with greater classroom misbehavior.

Regarding the psychometric properties of the model (shown in Table 5), the VIFs were acceptable (range 1.025–1.223). However, the residuals differed significantly from a normal distribution (Shapiro-Wilk = .947, $p < .001$). Nevertheless, the Shapiro-Wilk test is very sensitive to deviations from normality in large samples, and the statistics for skewness of 0.747 (standard error = 0.235) and kurtosis of -0.081 (standard error = 0.465) are within the acceptable range for normally distributed data [54].

Table 4
Linear regression analysis predicting grades in the high school sample.

	B	*SE B*	β	*t*	*p*
Age	−0.798	0.557	−0.136	−1.434	0.155
Vocabulary	−.011	.045	−.024	−.251	0.803
Initiation time	1.670	.859	.198	1.943	0.055
Suppression time	−1.370	0.455	−0.308	−3.011	0.003
Suppression errors	−0.025	0.047	−.052	−.536	0.593
Working memory	0.652	0.305	0.200	2.136	0.035

B = unstandardized beta, *SE B* = standard error for *B*, β = standardized beta.

Table 5
Linear regression analysis predicting misbehavior in the high school sample.

	B	*SE B*	β	*t*	*p*
Age	−1.881	0.848	−0.214	−2.217	0.029
Vocabulary	0.079	0.068	0.110	1.149	0.253
Initiation time	−1.524	1.310	−0.120	−1.163	0.248
Suppression time	0.167	0.694	.025	0.241	0.810
Suppression errors	−0.029	0.071	−.040	−0.407	0.685
Working memory	−1.136	.465	−0.233	−2.441	0.016

B = unstandardized beta, *SE B* = standard error for *B*, β = standardized beta.

4. General discussion

Since the case study of Phineas Gage, it has been known that damage to the frontal lobes often produces verbal disinhibition. John M. Harlow, the physician who treated Gage after his accident described him thus “He is fitful, irreverent, indulging at times in the grossest profanity (which was not previously his custom)...”, p 227 [23]. Such observations of verbal disinhibition have frequently been repeated in patients with damage to the frontal lobes [24–27]. Furthermore, brain imaging studies have consistently indicated that verbal response suppression involves activation mainly of the frontal lobes [55,56], and neuropsychological studies have shown that frontal damaged patients have particular difficulties with verbal response suppression tasks [20,28–30].

Although other neurocognitive functions associated with frontal lobe functioning have been widely investigated in non-clinical domains, such as education (particularly working memory), verbal response suppression has rarely been studied. In a previous study, we reported that the Hayling Test, a neuropsychological assessment designed to capture impairments in verbal response suppression after frontal lobe damage, was a particularly good predictor of academic grades of a mixed sample of undergraduate students studying a variety of majors [10]. This suggests that, like intelligence, it has a rather general association with academic performance, in contrast with working memory, which appears to be quite domain-specific, being particularly associated with language and mathematics performance.

Nevertheless, although intelligence test and the Hayling test performance both may have rather general associations with academic attainment, they appear to be independent associations. This is suggested by the fact that Hayling Test suppression scores do not correlate with intelligence test scores in healthy individuals [10,48]. Also, the neuropsychological evidence suggests that verbal response suppression on the Hayling Test is impaired by lesions to the frontal lobe, independently of concomitantly reduced general intelligence [29]. Finally, from an educational perspective, verbal response suppression scores on the Hayling Test and intelligence test scores explain independent amounts of variance in academic grades [10]. The current research attempted to explore in more detail the association between verbal response suppression and academic achievement, and several issues are discussed in turn below.

First, the original finding that Hayling Test Part 2 performance predicts academic achievement is replicated in two different samples. Furthermore, as previously observed [10], the relationship with academic achievement was numerically larger than for any of the other cognitive measures explored. In the undergraduate sample, greater suppression errors were associated with lower grades. In the high school sample, although there was no association with error scores, there was with response times: students with longer suppression response latencies had lower grades. Why two measures of response suppression are differentially associated depending on sample is unclear. However, it is possible that accuracy in response suppression and time for response suppression draw on different neurocognitive processes. This is suggested by a voxel-based lesion-symptom mapping study of patients with frontal lobe lesions which showed that accuracy on the Hayling Test Part 2 is dependent on the integrity of the orbitoventral prefrontal cortex while normal response latencies in the same task are dependent on the integrity of the medial-rostral prefrontal cortex [30]. An alternative explanation may simply be that the modifications made to the Hayling Test altered the way that participants traded-off speed and accuracy. It possible that on the standard task, participants focused primarily on accuracy, while in the modified version, participants prioritized fast responses or vice versa. Such a change in emphasis between speed and accuracy of performance could change whether it is response latencies or error scores which are associated with academic grades.

A second observation is that spontaneous strategy use was associated with academic performance. Again, this was shown in the two different samples. For the undergraduate sample, there was a positive correlation between strategy use and grades, probably driven mainly by the use of visual strategies, such as naming objects in view during assessment. For the high school sample, there was a negative correlation between the ostensibly semantic strategy of producing words related to previous responses. However, as discussed earlier, an alternative interpretation is that such semantically-related responses occur involuntarily due to priming, and in fact attention is required to prevent them. An implication of this is that for the Hayling Test in general, more investigation is needed to clarify whether both semantic and visual responses should be considered as strategies applied by participants. Nevertheless, the current results suggest that at least spontaneous visual strategy use is predictive of academic achievement of university students, and adds to our understanding of the role of top-down cognitive control in success within higher education.

A third observation is that strategy application is not the sole cause of the associations between Hayling Test Part 2 performance and academic grades. The modified Hayling Test, adapted specifically for this research, was able to significantly reduce the number of visual strategies used, yet an association remained between GPA and suppression performance. This suggests that the response suppression ability is associated with academic achievement, at least in the high school sample.

A fourth observation is that variation in neither working memory ability nor general verbal ability underlie the association between suppression ability and GPA. We argue this because, in general, measurements of working memory capacity (the Counting Span and the Reading Span Tests) did not significantly predict grades, whereas measures of response suppression did. Furthermore, in the regression model of scores on the modified Hayling Test, both suppression latencies and working memory scores were significantly associated with grades, suggesting independent amounts of variance in grades of the two different cognitive measures. Similarly, WAIS Vocabulary scores were not correlated with grades in the high school sample, and Hayling suppression latencies were significant even with WISC Vocabulary scores entered into the model.

A fifth observation is related to the fact that even though there were no significant correlations between grades and Counting Span (in the undergraduate sample) or Reading Span performance (in the high school sample), there was a significant correlation with the working memory test embedded within the modified Hayling Test (used only with the high school sample). In general, this is consistent with previous research that has suggested a role for working memory in high school academic achievement [57], but not at university level [10]. Furthermore, even at high school level, whether or not an association is observed between working memory scores and grades is dependent on the demands of the specific working memory task employed [57].

The final observation refers to the links to classroom misbehavior. Given that response suppression appears to have associations with academic achievement, it is tempting to hypothesize that it will also be linked to classroom misbehavior. One could see how poor response suppression could result in the types of misbehavior we recorded, e.g. frequent interruptions or aggression. Indeed, one of the key predictors of classroom misbehavior is a diagnosis of attention-deficit/hyperactivity disorder (ADHD) [58], and a prominent theory of the neuropsychological profile of ADHD is that behavioral inhibition is a core deficit underlying the problems with other executive processes [59]. However, we could find no links between response suppression (or for that matter strategy use) as measured by the Hayling Test and classroom misbehavior.

Although with the Reading Span Test of working memory there was no link to classroom misbehavior, on the embedded working memory span task in the modified Hayling Test, such an association was detected. Poor working memory scores predicted higher ratings of classroom misbehavior by teachers. The reason that this was only detected in our embedded working memory task may be because that version involved different cognitive demands. In our memory task embedded within the modified Hayling Test, the requirement was simply to recall a single word, though while performing several other attention-demanding tasks between presentation and recall. Therefore, it likely challenges central resources (such as the central executive) more than phonological storage capacity. Our finding is consistent with research with younger children that has suggested that although working memory scores predict classroom misbehavior, it is variation in processing efficiency, not working memory capacity *per se* that drives the relationship [60].

In addition, although we did not measure ADHD, it may be that our Misbehavior Scale was capturing information about ADHD symptomology. This could explain why high misbehavior scores were statistically linked to relatively poorer working memory performance: Children with ADHD typically show poor working memory, this involves phonological and visuospatial processes, but central executive processing efficiency is particularly affected [61]. This would be consistent with our finding that the association was only detected in the embedded version of the test, assumed to particularly challenge central attentional processes. Why response suppression efficiency is not associated with classroom behavior is less clear, particularly if we assume that misbehavior is related to ADHD-type symptomology and that a response inhibition deficit is the core cognitive problem in ADHD [59]. However, other theories of ADHD do not emphasize response suppression and instead point to working memory problems as being central to the cognitive profile of the disorder, proposing that a core working memory deficit that leads to disorganized behavior such as stimulation seeking and hyperactivity [61,62]. The current results could be interpreted in a similar fashion, as they suggest that relative weaknesses in working memory are associated with greater classroom misbehavior.

In summary, therefore, our results suggest the Hayling Test of verbal response suppression may be useful for identifying the neurocognitive processes associated with real-life success, such as in education. In particular, we presented evidence that verbal response suppression ability and spontaneous strategy use are both related to academic achievement and that working memory ability is related to classroom misbehavior. The substantial evidence pointing to the role of the prefrontal cortex in Hayling Task performance [20,28,30,55,56], and the fact that that region is very late to mature, associated with lengthy development of response suppression ability [63,64], may suggest that, in young people, efficiency of processing of the prefrontal cortex underlies the associations reported here.

A further issue is the role of functional differentiation of cognitive function over the course of development. In children, there is a greater correlation between abilities such as working memory and response inhibition than there is in adolescence and adulthood [65]. This differentiation of functions as people move into adolescence and adulthood is mirrored by functional neuroanatomical changes in which spatially near brain regions become more functionally segregated, i.e. their activity levels become less correlated [66]. In particular, development from childhood to adolescence is associated with increased executive control ability, which is linked to increasing focal activation of prefrontal cortical regions [67].

That developing into adulthood is associated with greater cognitive and neural differentiation may explain why, in children, intelligence is linked to academic achievement, while in older adolescents such as reported here, more specific executive processes such as response suppression appear to be important factors. This differentiation may also explain why Hayling Test performance differed somewhat between our high-school and university samples. Both groups performed a version of the standard Hayling Test, but it was only in the older university students that response suppression error scores were associated with grades. It is possible that as the prefrontal cortex and specific executive processes become more differentiated, their associated impacts on variation in real-life performance such as educational grades become more apparent. Nevertheless, one has to be extremely cautious linking behavioral data such as this to anatomy, and these tentative hypotheses should be evaluated accordingly. Although, some support comes from an fMRI study with medical students that linked academic achievement to level of prefrontal activation [68].

However, some limitations of the current research should be acknowledged. The research is observational, and no causation between task performance and academic grades can be asserted. In addition, the Hayling Test, which was the central assessment, is primarily a clinical test [21], and may not measure the same construct in healthy young people as it does in neurological patients. Consequently, whereas a neuropsychologist may infer frontal lobe involvement from Hayling Task performance in the clinic, the same inference may be less valid in healthy participants. Nevertheless, a study of Hayling Test performance over the healthy adult lifespan suggested that the changes to performance appear to be consistent with age-related changes to the frontal lobes, and the performance of the oldest healthy participants was equivalent to that of younger frontal lesioned patients [69]. This suggests that the processes underlying Hayling Test performance in healthy and neurologically impaired patients may be equivalent.

A related issue is that due to floor and ceiling effects in healthy participants, the Hayling Test may fail to adequately measure performance as a continuous variable. On the other hand, our modified Hayling Test was able to raise response times and error rates, and the extra demands may make the test more useful in non-clinical samples than the standard clinical scale. Further research may be able to resolve these issues and shed further light on why verbal response suppression seems to be a particularly good predictor of academic achievement of late adolescent and adult students.

Ethical statement

The research reported was reviewed and approved by the Comite de Bioetica at Universidad San Francisco de Quito. This is an officially recognised research ethics committee.

All participants provided written informed consent.

The research was conducted in accordance with the Declaration of Helsinki.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Declaration of Competing Interest

None.

Acknowledgment

We would like to thank Marco Antonio Lopez for assistance with data processing.

References

[1] J. Hattie, The applicability of visible learning to higher education, Scholarsh. Teach. Learn. Psychol. 1 (1) (2015) 79–91, https://doi.org/10.1037/stl0000021.
[2] T.P. Alloway, R.G. Alloway, Investigating the predictive roles of working memory and IQ in academic attainment, J. Exp. Child Psychol. 106 (1) (2010) 20–29, https://doi.org/10.1016/j.jecp.2009.11.003.
[3] I.J. Deary, S. Strand, P. Smith, C. Fernandes, Intelligence and educational achievement, Intell 35 (1) (2007) 13–21, https://doi.org/10.1016/j.intell.2006.02.001.
[4] M. Arya, S.P. Maurya, Relationship between creativity, intelligence and academic achievement among school going children, Stud. Home Community Sci. 10 (1) (2016) 1–7, https://doi.org/10.1080/09737189.2016.11885359.
[5] A. Diseth, The relationship between intelligence, approaches to learning and academic achievement, Scand. J. Educ. Res. 46 (2) (2002) 219–230, https://doi.org/10.1080/00313830220142218.
[6] A. Furnham, T. Chamorro-Premuzic, F. McDougall, Personality, cognitive ability, and beliefs about intelligence as predictors of academic performance, Learn. Individ. Differ. 14 (2003) 49–66, https://doi.org/10.1016/j.lindif.2003.08.002.
[7] K.P. Raghubar, M.A. Barnes, S.A. Hecht, Working memory and mathematics: A review of developmental, individual difference, and cognitive approaches, Learn. Individ. Differ. 20 (2) (2010) 110–122, https://doi.org/10.1016/j.lindif.2009.10.005.
[8] H.L. St Clair-Thompson, S.E. Gathercole, Executive functions and achievements in school: shifting, updating, inhibition, and working memory, Q. J. Exp. Psychol. 59 (4) (2006) 745–759, https://doi.org/10.1080/17470210500162854.
[9] D. Giofrè, I.C. Mammarella, C. Cornoldi, The structure of working memory and how it relates to intelligence in children, Intell. 41 (5) (2013) 396–406, https://doi.org/10.1016/j.intell.2013.06.006.
[10] G. Pluck, C.B. Ruales-Chieruzzi, E.J. Paucar-Guerra, M.V. Andrade-Guimaraes, A.F. Trueba, Separate contributions of general intelligence and right prefrontal neurocognitive functions to academic achievement at university level, Trends Neurosci. Educ. 5 (4) (2016) 178–185, https://doi.org/10.1016/j.tine.2016.07.002.
[11] R.G. Brown, G. Pluck, Negative symptoms: the 'pathology' of motivation and goal-directed behaviour, Trends Neurosci. 23 (9) (2000) 412–417, https://doi.org/10.1016/S0166-2236(00)01626-X.
[12] G. Pluck, K.H. Lee, Negative symptoms and related disorders of diminished goal directed behavior, Minerva Psichiatr. 54 (1) (2013) 15–29.
[13] R.J. Blair, L. Cipolotti, Impaired social response reversal. A case of 'acquired sociopathy', Brain 123 (6) (2000) 1122–1141, https://doi.org/10.1093/brain/123.6.1122.
[14] M.A. Cato, D.C. Delis, T.J. Abildskov, E. Bigler, Assessing the elusive cognitive deficits associated with ventromedial prefrontal damage: a case of a modern-day Phineas Gage, J. Int. Neuropsychol. Soc. 10 (3) (2004) 453–465, https://doi.org/10.1017/S1355617704103123.
[15] P.J. Eslinger, A.R. Damasio, Severe disturbance of higher cognition after bilateral frontal lobe ablation: patient EVR, Neurology 35 (12) (1985) 1731–1741, https://doi.org/10.1212/WNL.35.12.1731.
[16] E.T. Heck, J.B. Bryer, Superior sorting and categorizing ability in a case of bilateral frontal atrophy: an exception to the rule, J. Clin. Exp. Neuropsychol. 8 (3) (1986) 313–316, https://doi.org/10.1080/01688638608401321.
[17] T. Shallice, P.W. Burgess, Deficits in strategy application following frontal lobe damage in man, Brain 114 (2) (1991) 727–741, https://doi.org/10.1093/brain/114.2.727.
[18] A. Miyake, N.P. Friedman, The nature and organization of individual differences in executive functions: four general conclusions, Curr. Dir. Psychol. Sci. 21 (1) (2012) 8–14, https://doi.org/10.1177/0963721411429458.
[19] L. Cipolotti, B. Spano, C. Healy, C. Tudor-Sfetea, E. Chan, M. White, F. Biondo, J. Duncan, T. Shallice, M. Bozzali, Inhibition processes are dissociable and lateralized in human prefrontal cortex, Neuropsychologia 93 (2016) 1–12, https://doi.org/10.1016/j.neuropsychologia.2016.09.018.
[20] P.W. Burgess, T. Shallice, Response suppression, initiation and strategy use following frontal lobe lesions, Neuropsychologia 34 (4) (1996) 263–272, https://doi.org/10.1016/0028-3932(95)00104-2.
[21] P.W. Burgess, T. Shallice, Hayling and Brixton Tests: Manual, Pearson Assessment, Oxford, UK, 1997.
[22] J.D. Duffy, J.J. Campbell 3rd, The regional prefrontal syndromes: a theoretical and clinical overview, J. Neuropsychiatry Clin. Neurosci. 6 (4) (1994) 379–387, https://doi.org/10.1176/jnp.6.4.379.
[23] J.M. Harlow, Recovery from the passage of an iron bar through the head, Hist. Psychiatry 4 (14) (1993) 274–281, https://doi.org/10.1177/0957154X9300401407 1868.
[24] N. Kapur, P. Hutchinson, E. Berry, K. Hawkins, D. Llewellyn, B. Wilson, Executive dysfunction in a case of transoral-frontal self-inflicted gunshot injury, Brain Inj. 23 (12) (2009) 985–989, https://doi.org/10.3109/02699050903373469.
[25] L.C. Keegan, H. McAdam, Swearing after traumatic brain injury: A linguistic analysis, J. Int. Res. Commun. Disord. 7 (1) (2016) 101–122, https://doi.org/10.1558/jircd.v7i1.21570.

[26] B.L. Miller, L. Chang, I. Mena, K. Boone, I.M. Lesser, Progressive right fronto-temporal degeneration: clinical, neuropsychological and SPECT characteristics, Dement 4 (3-4) (1993) 204–213, https://doi.org/10.1159/000107324.

[27] N. Nagaratnam, P. Bou-Haidar, H. Leung, Confused and disturbed behaviour in the elderly following silent frontal lobe infarction, Am. J. Alzheimers Dis. Other Dement. 18 (6) (2003) 333–339, https://doi.org/10.1177/153331750301800608.

[28] G.A. Robinson, L. Cipolotti, D.G. Walker, V. Biggs, M. Bozzali, T. Shallice, Verbal suppression and strategy use: a role for the right lateral prefrontal cortex? Brain 138 (4) (2015) 1084–1096, https://doi.org/10.1093/brain/awv003.

[29] M. Roca, A. Parr, R. Thompson, A. Woolgar, T. Torralva, N. Antoun, F. Manes, J. Duncan, Executive function and fluid intelligence after frontal lobe lesions, Brain 133 (1) (2010) 234–247, https://doi.org/10.1093/brain/awp269.

[30] E. Volle, A. de Lacy Costello, L.M. Coates, C. McGuire, K. Towgood, S. Gilbert, S. Kinkingnehun, J.E. McNeil, R. Greenwood, B. Papps, M. van den Broeck, P.W Burgess, Dissociation between verbal response initiation and suppression after prefrontal lesions, Cereb. Cortex 22 (10) (2012) 2428–2440, https://doi.org/10.1093/cercor/bhr322.

[31] R.M. Roth, P.K. Isquith, G.A. Gioia, Behavior Rating Inventory of Executive Function - Adult Version (BRIEF-A), Psychological Assessment Resources, Lutz, FL, 1996.

[32] O. Orlando Reynoso Orozco, T.E. Méndez-Luévano, Can academic achievement be predicted? Behavior regulation as an indicator of academic achievement in higher education students, Diálogos Sobre Educ. 16 (2018) 1–19, https://doi.org/10.32870/dse.v0i16.397.

[33] J.M. Strayhorn Jr., Self-control: theory and research, J. Am. Acad. Child Adolesc. Psychiatry 41 (1) (2002) 7–16, https://doi.org/10.1097/00004583-200201000-00006.

[34] A.L. Duckworth, M.E. Seligman, Self-discipline outdoes IQ in predicting academic performance of adolescents, Psychol. Sci. 16 (12) (2005) 939–944, https://doi.org/10.1111/j.1467-9280.2005.01641.x.

[35] P.A. Gouveia, S.M. Brucki, S.M. Malheiros, O.F. Bueno, Disorders in planning and strategy application in frontal lobe lesion patients, Brain Cogn 63 (3) (2007) 240–246, https://doi.org/10.1016/j.bandc.2006.09.001.

[36] I. Obeso, L. Wilkinson, E. Casabona, M.L. Bringas, M. Alvarez, L. Alvarez, N. Pavon, M.C. Rodriguez-Oroz, R. Macias, J.A. Obeso, M. Jahanshahi, Deficits in inhibitory control and conflict resolution on cognitive and motor tasks in Parkinson's disease, Exp. Brain Res. 212 (3) (2011) 371–384, https://doi.org/10.1007/s00221-011-2736-6.

[37] A.R. Conway, M.J. Kane, M.F. Bunting, D.Z. Hambrick, O. Wilhelm, R.W. Engle, Working memory span tasks: A methodological review and user's guide, Psychon. Bull. Rev. 12 (5) (2005) 769–786, https://doi.org/10.3758/BF03196772.

[38] D.C. Delis, E. Kaplan, J.H. Kramer, Delis-Kaplan Executive Function System: Technical Manual, Psychological Corporation, San Antonio, TX, 2001.

[39] K. Arbuthnott, J. Frank, Trail Making Test, Part B as a measure of executive control: validation using a set-switching paradigm, J. Clin. Exp. Neuropsychol. 22 (4) (2000) 518–528, https://doi.org/10.1076/1380-3395(200008)22:4;1-0;FT518.

[40] A.J. Bishara, J.B. Hittner, Testing the significance of a correlation with nonnormal data: comparison of Pearson, Spearman, transformation, and resampling approaches, Psychol. Methods 17 (3) (2012) 399–417, https://doi.org/10.1037/a0028087.

[41] R. Bull, K.A. Espy, S.A. Wiebe, Short-term memory, working memory, and executive functioning in preschoolers: longitudinal predictors of mathematical achievement at age 7 years, Dev. Neuropsychol. 33 (3) (2008) 205–228, https://doi.org/10.1080/87565640801982312.

[42] M. Daneman, P.A. Carpenter, Individual differences in working memory and reading, J. Verbal Learn. Verbal Behav. 19 (4) (1980) 450–466, https://doi.org/10.1016/S0022-5371(80)90312-6.

[43] T. Rutherford, M. Buschkuehl, S.M. Jaeggi, G. Farkas, Links between achievement, executive functions, and self-regulated learning, Appl. Cogn. Psychol. 32 (6) (2018) 763–774, https://doi.org/10.1002/acp.3462.

[44] V. Stenback, M. Hallgren, B. Lyxell, B. Larsby, The Swedish Hayling task, and its relation to working memory, verbal ability, and speech-recognition-in-noise, Scand. J. Psychol. 56 (3) (2015) 264–272, https://doi.org/10.1111/sjop.12206.

[45] D. Villagómez, G. Pluck, P. Almeida, Relación entre la memoria de trabajo, inhibición de respuesta, y habilidad verbal con el éxito académico y el comportamiento en adolescente, Maskana 8 (2017) 87–100 (Actos del Simposio Internacional de Neurociencias: Interacción Humana con la Inteligencia Artificial, la Realidad Virtual y el Medio Ambiente, 2017).

[46] C.K. Conners, Conners rating scale—Revised, Multihealth Systems, (2001) North Tunawanda, NY.

[47] A. Perez-Perez, J.A. Matias-Guiu, I. Caceres-Guillen, T. Rognoni, M. Valles-Salgado, M. Fernandez-Matarrubia, T. Moreno-Ramos, J. Matias-Guiu, The Hayling Test: development and normalization of the Spanish version, Arch. Clin. Neuropsychol. 31 (5) (2016) 411–419, https://doi.org/10.1093/arclin/acw027.

[48] K. Stedal, M. Rose, R. Jonsson, L. Harvey, B. Lask, Female adolescents' performance on the Hayling Test, Psychol. Neurosci. 9 (2) (2016) 255–266, https://doi.org/10.1037/pne0000056.

[49] D. Wechsler, Wechsler Intelligence Scale for Children, Fourth Edition-Spanish, PsychCorp, Madrid, Spain, 2004.

[50] M.R. Elosúa, F. Gutiérrez, J.A.G. Madruga, J.L. Luque, M. Gárate, Adaptación española del «Reading Span Test» de Daneman y Carpenter, Psicothema 8 (2) (1996) 383–395.

[51] L.A. Clark, D. Watson, Constructing validity: Basic issues in objective scale development, Psychol. Assess. 7 (3) (1995) 309–319, https://doi.org/10.1037/1040-3590.7.3.309.

[52] T. Heyman, B. Van Rensbergen, G. Storms, K.A. Hutchison, S. De Deyne, The influence of working memory load on semantic priming, J. Exp. Psychol. Learn. Mem. Cogn. 41 (3) (2015) 911–920, https://doi.org/10.1037/xlm0000050.

[53] A. Vallesi, Organisation of executive functions: hemispheric asymmetries, J. Cogn. Psychol. 24 (4) (2012) 367–386, https://doi.org/10.1080/20445911.2012.678992.

[54] H.Y. Kim, Statistical notes for clinical researchers: assessing normal distribution (2) using skewness and kurtosis, Restor. Dent. Endod. 38 (1) (2013) 52–54, https://doi.org/10.5395/rde.2013.38.1.52.

[55] G.I. de Zubicaray, F.O. Zelaya, C. Andrew, S.C. Williams, E.T. Bullmore, Cerebral regions associated with verbal response initiation, suppression and strategy use, Neuropsychologia 38 (9) (2000) 1292–1304, https://doi.org/10.1016/S0028-3932(00)00026-9.

[56] D.A. Nathaniel-James, P. Fletcher, C.D. Frith, The functional anatomy of verbal initiation and suppression using the Hayling Test, Neuropsychologia 35 (4) (1997) 559–566, https://doi.org/10.1016/S0028-3932(96)00104-2.

[57] K. Kanerva, V. Kalakoski, The predictive utility of a working memory span task depends on processing demand and the cognitive task, Appl. Cogn. Psychol. 30 (5) (2016) 681–690, https://doi.org/10.1002/acp.3243.

[58] G.J. August, G.M. Realmuto, A.W. MacDonald 3rd, S.M. Nugent, R. Crosby, Prevalence of ADHD and comorbid disorders among elementary school children screened for disruptive behavior, J. Abnorm. Child Psychol. 24 (5) (1996) 571–595, https://doi.org/10.1007/BF01670101.

[59] R.A. Barkley, Behavioral inhibition, sustained attention, and executive functions: constructing a unifying theory of ADHD, Psychol. Bull. 121 (1) (1997) 65–94, https://doi.org/10.1037/0033-2909.121.1.65.

[60] C. Jarrold, N. Mackett, D. Hall, Individual differences in processing speed mediate a relationship between working memory and children's classroom behaviour, Learn. Individ. Differ. 30 (2014) 92–97, https://doi.org/10.1016/j.lindif.2013.10.016.

[61] M.D. Rapport, R.M. Alderson, M.J. Kofler, D.E. Sarver, J. Bolden, V. Sims, Working memory deficits in boys with attention-deficit/hyperactivity disorder (ADHD): the contribution of central executive and subsystem processes, J. Abnorm. Child Psychol. 36 (6) (2008) 825–837, https://doi.org/10.1007/s10802-008-9215-y.

[62] M.D. Rapport, K.M. Chung, G. Shore, P. Isaacs, A conceptual model of child psychopathology: implications for understanding attention deficit hyperactivity disorder and treatment efficacy, J. Clin. Child Psychol. 30 (1) (2001) 48–58, https://doi.org/10.1207/S15374424JCCP3001_6.

[63] E.R. Sowell, P.M. Thompson, C.J. Holmes, T.L. Jernigan, A.W. Toga, In vivo evidence for post-adolescent brain maturation in frontal and striatal regions, Nat. Neurosci. 2 (10) (1999) 859–861, https://doi.org/10.1038/13154.

[64] X. Zhou, D. Zhu, S.G. King, C.J. Lees, A.J. Bennett, E. Salinas, T.R. Stanford, C. Constantinidis, Behavioral response inhibition and maturation of goal representation in prefrontal cortex after puberty, Proc. Natl. Acad. Sci. USA (2016), https://doi.org/10.1073/pnas.1518147113.

[65] Y.L. Shing, U. Lindenberger, A. Diamond, S.C. Li, M.C. Davidson, Memory maintenance and inhibitory control differentiate from early childhood to adolescence, Dev. Neuropsychol. 35 (6) (2010) 679–697, https://doi.org/10.1080/87565641.2010.508546.

[66] D.A. Fair, A.L. Cohen, J.D. Power, N.U. Dosenbach, J.A. Church, F.M. Miezin, B.L. Schlaggar, S.E. Petersen, Functional brain networks develop from a "local to distributed" organization, PLoS Comput. Biol. 5 (5) (2009) e1000381, , https://doi.org/10.1371/journal.pcbi.1000381.

[67] D. Yurgelun-Todd, Emotional and cognitive changes during adolescence, Curr. Opin. Neurobiol. 17 (2) (2007) 251–257, https://doi.org/10.1016/j.conb.2007.03.009.

[68] K. Veroude, J. Jolles, M. Knežević, C.M.P. Vos, G. Croiset, L. Krabbendam, Anterior cingulate activation during cognitive control relates to academic performance in medical students, Trends. Neurosci. Educ. 2 (2013) 100–106, https://doi.org/10.1016/j.tine.2013.10.001.

[69] E.C. Gibson, M.S. Barker, A.K. Martin, G.A. Robinson, Initiation, inhibition and strategy generation across the healthy adult lifespan, Arch. Clin. Neuropsychol. 34 (4) (2019) 511–523, https://doi.org/10.1093/arclin/acy057.