Cognition, Brain, Behavior. An Interdisciplinary Journal

ISSN: Print 2247-9228, Online 2601-226X
Volume XXII, No. 3 (September), 175-196
doi:10.24193/cbb.2018.22.12

COGNITION BRAIN BEHAVIOR
AN INTERDISCIPLINARY JOURNAL

Lexical reading ability predicts academic achievement at university level

Graham Pluck[*]

Quito Brain and Behavior Lab, Universidad San Francisco de Quito, Quito, Ecuador

Abstract

The prediction of academic achievement with cognitive testing has important practical and theoretical implications. Although there are associations between cognitive performance and school grades, this does not extend well to higher education. An alternative to intelligence or IQ testing as a predictor of grades is lexical reading. Access to the mental lexicon is particularly important when reading aloud forms with irregular grapheme-phoneme conversion or unpredictable stress patterns. Such words have to be learnt and already present in declarative lexical stores to be reliably pronounced correctly. A sample of 102 participants was assessed with a standard test of lexical reading, the Word Accentuation Test (WAT) which involves pronunciation of a list of low-frequency words. To assess semantic-conceptual contributions to performance, they also completed a version of the WAT in which they read the same words embedded in sentences, as well as performed a stem-completion naming task. All the tests had good psychometric properties, and the two tests of lexical reading were both significant predictors of academic achievement. Reading aloud in sentences did improve pronunciation accuracy, particularly for poor readers. However, it did not add anything to the ability to predict GPA. In fact, when using linear regression, the best predictor of academic achievement (R = .40) was the pronunciation of low-frequency word forms presented in isolation. This predictive ability suggests that the breadth of learnt stores of lexical forms, but not associated semantic processing, may be an individual difference that can potentially predict college-level academic achievement. The association with dyslexia is also considered.

Keywords: *reading, pronunciation, academic achievement, surface dyslexia, mental lexicon*

The prediction of academic achievement with cognitive assessments has long been a concern of psychologists. In fact, the very concept of the intelligence test was a consequence of a need to identify academically weak students for special educational provision (Binet, 1903). This historical approach no doubt has some

[*]Corresponding author:
E-mail: g.c.pluck@gmail.com

practical applications, even in the twenty-first century, where the emphasis has shifted, rightly or wrongly, to identifying academically strong students to promote expertise and elite education (Rasmussen & Lingard, 2018).

However, a quite different reason for understanding the relationships between cognitive performance and academic achievement is for what it can inform about cognitive systems. This is because, for example, knowing that process *x* predicts certain forms of success (or failure) provides a form of validity to the theory of process *x*, as well as it informs about the biological function of the process. This can be seen clearly in the theory of working memory (Baddeley, 2017; Baddeley & Hitch, 1974). This suggests a tripartite system of active short-term memory composed of a limited resource and general central executive, as well as slave systems of a visuospatial sketchpad and a phonological loop. At the time the working memory model was first presented, although it was able to explain a range of experimental observations, it was not clear what biological function the components such as the phonological loop actually served (Baddeley, 2003).

It had been assumed that working memory played some part in the normal processing of language, but that this might be limited to decoding long-sentences that have complex syntax. However, it was a classic study with undergraduate participants by Daneman and Carpenter (1980), which unequivocally demonstrated the crucial role of phonological working memory in language comprehension, a result that has been frequently replicated (e.g., Iglesias-Sarmiento, López, & Rodríguez, 2015). By using a working memory span task, Daneman and Carpenter (1980) demonstrated that individual differences in phonological working memory are highly correlated with verbal SAT scores, as well as experimental measures of information recall and pronominal referencing from read texts. The authors also showed that the same phenomenon occurs with auditory language comprehension. It is now clear from studies with school children that phonological working memory is related to achievement in language studies, while both phonological and visuospatial working memory have links to mathematics achievement (Formoso et al., 2018; St Clair-Thompson & Gathercole, 2006). These essentially provide evidence of the functional importance of working memory.

Cognitive prediction in higher education

Although such associations between cognitive systems and achievement are now well recognized in school-aged children, the situation is much less clear in higher education. As university level study is at the apex of modern educational systems, it might be thought that the contribution of individual differences in cognitive abilities to achievement would be at its strongest. However, studies of intelligence testing suggest that it contributes very little to the performance of university students. In fact, one important meta-analysis of predictive factors in higher education found that the mean correlation of intelligence test scores with grades was only about .20

(Richardson, Abraham, & Bond, 2012), essentially a negligible association in terms of predictive ability.

Part of the reason for this is that intelligence is a very general, and poorly defined concept. There has been some advancement in examining the role of more specific cognitive systems, such as executive functions in higher education success. This has revealed that working memory ability appears to be of little importance, while good response suppression, perhaps because it allows students to focus on the educational material, is related to academic achievement (Pluck, Ruales-Chieruzzi, Paucar-Guerra, Andrade-Guimaraes, & Trueba, 2016). Another study used a composite score of a large set of executive function tests linked to prefrontal function and found significant correlations with university grades (Higgins, Peterson, Pihl, & Lee, 2007). Memory systems also appear to be important, with some evidence suggesting that the top-down modulation of non-declarative and declarative learning is particularly important for success in higher education (Pluck et al., in press).

The role of the language system in high-level goal-directed behavior

One alternative line of research is not on domain-general behavioral control mechanisms such as top-down executive control, but on the capacity of the language system. This is particularly pertinent considering recent theory, driven by neuroimaging findings, that postulates two fundamental systems in the human brain. One of these is a domain-general system that underlies fluid intelligence and many executive functions, based on parts of the frontal and parietal cortices. The other is a non-overlapping domain-specific language system that is active only for language processing, based on the temporal lobe and parts of the frontal cortices (Blank & Fedorenko, 2017; Campbell & Tyler, 2018; Woolgar, Duncan, Manes, & Fedorenko, 2018). It should be noted that language processing will often involve the domain-general system (for example verbal working memory), however, the parts of the language system that are domain-specific are never active for non-linguistic tasks. This is why they are considered separate systems.

That individual differences within the capacity of the language system might predict high-level academic achievement, provides a viable alternative to the attempts to predict performance from domain-general cognitive processes such as executive functions. Is there evidence that the language system may be more than communication, but actually involved in high-level control of intelligent goal-directed behavior? Looking at the language system from an evolutionary perspective, humans are clearly much more cognitively competent than all other animals (Premack, 2007). Human success has been to a large extent due to the development of brain regions allowing for tool use, which appears to be lateralized mainly to the left hemisphere (Lewis, 2006), as is the language system (Pinel & Dehaene, 2010). In fact, tool-use and language may have co-evolved in humans providing the basis for human higher cognitive ability (Vingerhoets et al., 2013). In support of this, using modern tool-making simulations, the development of primitive tool use has been linked to processing in the left-hemisphere lateralized Broca's area, usually associated with language ability (Stout & Chaminade, 2007). This suggests a role for the language system in high-level adaptive behavior.

Obviously, language is essential for various forms of verbal reasoning, additionally though, neuropsychological evidence suggests that left-hemisphere language systems are also essential for non-verbal reasoning (Baldo, Bunge, Wilson, & Dronkers, 2010). It is probably no coincidence that recent neuropsychological and functional imaging studies have identified a disproportionate role of the human left hemisphere in intelligence (Barbey et al., 2012; Santarnecchi, Emmendorfer, & Pascual-Leone, 2017). In summary, individual differences in capacity to use language likely underpin many forms of complex thought, and for this reason may contribute to real-life success in highly challenging situations, such as higher education.

Although the core language system may be independent of the domain-general processes underlying fluid intelligence and many executive functions, as described above, it is clearly linked to working memory, particularly the phonological loop. Furthermore, the contents of language are unquestionably learnt, and in this sense, language must share some commonality with memory systems. The most basic distinction in human memory is between declarative and procedural memory, with declarative being linked to the temporal lobe and procedural to the basal ganglia (Squire & Dede, 2015). It appears that different aspects of language processing rely on declarative and procedural learning, in particular lexical word forms may be part of the declarative system, while grammatical rules may be part of the procedural system. Evidence for this comes from studies of the correlations between memory skills and language skills in children. These have shown that lexical ability correlates significantly with measures of declarative learning ability, but not with procedural learning ability, while grammatical ability does not correlate with declarative learning ability, but it does correlate with procedural learning ability (Kidd, 2012; Lum, Conti-Ramsden, Page, & Ullman, 2012).

There is further support for this distinction from neuropsychology, where it has been demonstrated that patients with impaired declarative memory caused by temporal lobe damage tend to make errors with irregular spelt words (e.g. making the past tense of the English word 'dig' → 'dug'.) In contrast, patients with impaired procedural memory caused by basal ganglia disease tend to make errors with regular word forms (e.g. making the past tense of the English word 'cook' → 'cooked'.) Together this evidence suggests that lexical word forms, particularly irregular forms, are learnt declaratively and grammar is at least stored procedurally (Ullman, 2013).

The current research focuses on the role of declarative language processing as a potential predictor of academic achievement in higher education. This is because there is considerable evidence that reading of irregular word forms is highly correlated with psychometrically measured intelligence (Alves, Simoes, & Martins, 2012; Bright, Hale, Gooch, Myhill, & van der Linde, 2018; Wechsler, 2011). On the other hand, individual differences in procedural learning ability do not appear to be linked to intelligence (Kaufman et al., 2010). In addition to reading, naming, which by definition is declarative, is also highly associated with intelligence (Storms, Saerens, & De Deyn, 2004).

Furthermore, both naming and reading irregular word forms share common stages of cognitive processing, in particular the phonological output lexicon stage, which applies information about the metrical structure of the word to be uttered, such as the number of syllables, and the stress pattern (Gvion & Friedmann, 2016). The phonological output lexicon is part of what is known as the lexical route to word reading. It is necessary for reading of word forms that are non-regular (regular spelt or stressed forms can also be read through a sub-lexical route that relies on grapheme-phoneme conversion). Only basic word stems are thought to be processed here, with suffixes added later, confirming that the phonological output buffer is part of the declarative system (Ullman, 2013). It is also sensitive to word frequency, with low-frequency words more difficult to access.

A key difference though between reading irregular forms through the lexical route and visual naming, is that the latter must be processed through the semantic system, this is because visual categorical recognition systems are needed (for example to recognize an image of an asteroid and then proceed to name it). However, for lexical reading, although semantic access may assist processing, it is not essential. Evidence for the functional independence of the lexical route in word reading comes from a double dissociation between naming and lexical reading. Patients with Alzheimer's disease perform poorly on semantic language tasks, but nevertheless have preserved ability to read irregular word forms (Mardh, Nagga, & Samuelsson, 2013), while some cases of dyslexia have impaired lexical reading (surface dyslexia) in the presence of normal semantic processing (Peterson, Pennington, Olson, & Wadsworth, 2014). Therefore, the correct pronunciation of an irregular word can be made without semantic access, providing the word is known and the idiosyncratic metric structure is stored in the phonological output lexicon. A summary of this model, showing how naming proceeds through the conceptual system, while word reading can be achieved by either the lexical route or through a grapheme-phoneme conversion route is shown in Figure 1, adapted from Gvion and Friedmann (2016).

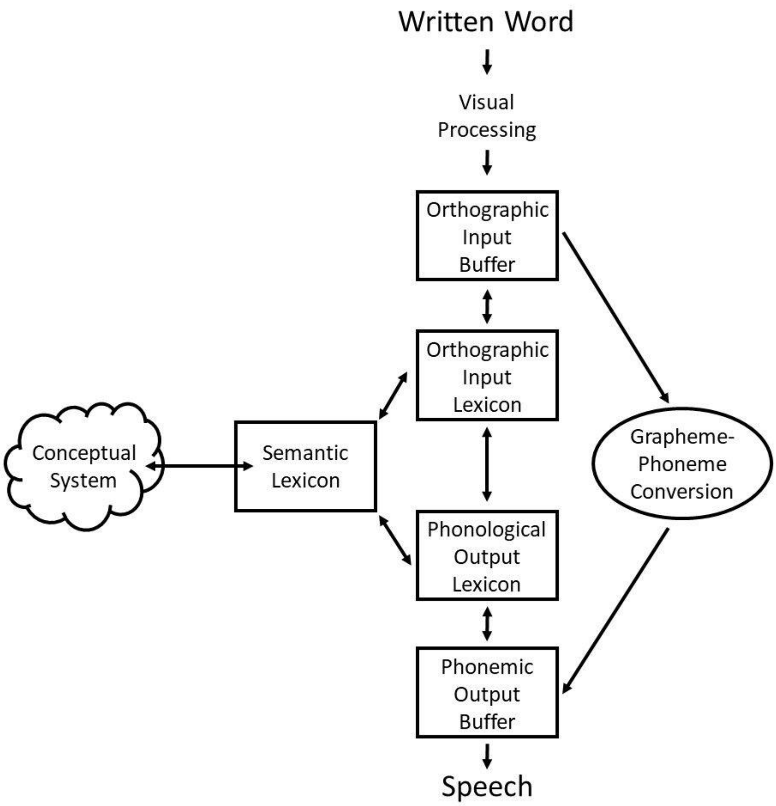

Figure 1. A cognitive model or speech production showing the lexical and sub-lexical routes to reading, adapted from Gvion and Friedmann (2016)

Current study

In summary, two key aspects of the declarative language system, which appear closely linked to intelligence, and may be candidates for predicting real-life academic achievement, are lexical word reading and visual naming. Particularly if they involve the production of low-frequency word forms. Although testing these abilities is rather similar - participants either name objects or read words aloud - the tasks differ on the extent to which they rely on categorical processing.

To investigate how these processes may predict real-life academic performance, a standardized word reading task was employed with Spanish-speaking university students. This task, the Word Accentuation Test (Del Ser, Gonzalez-Montalvo, Martinez-Espinosa, Delgado-Villapalos, & Bermejo, 1997; Gomar et al., 2011), uses reading aloud of low-frequency idiosyncratic words to force participants to read through the lexical route. This therefore tests their access to low-frequency lexical forms. A custom-made version of the same test was used to enhance performance by providing semantic context (reading the words in sentences). Finally, a naming task of images of low-frequency words was employed, in which multiple cues are provided to challenge the route to speech through the categorical system. In this way, it was hypothesized that the pronunciation of irregular word forms will be associated with academic achievement of university students. This hypothesis would be true if the breadth of entries in the lexical reading route is associated with academic achievement (either as a cause or consequence). Alternatively, it was hypothesized that the tests in which semantic aspects are employed, the sentence version of the Word Accentuation Test, and particularly the naming task, will be associated with academic achievement. This would be true if conceptual-semantic knowledge is particularly important in academic achievement, more so than simply the breadth of entries in the lexical reading route stores.

METHOD

Participants

One-hundred and two adult participants were recruited at Universidad San Francisco de Quito in Ecuador. The majority of these were undergraduate students (n = 86, 84.90%) from a range of different study majors, the remaining 16 participants were employees of the University (e.g. cleaners, security guards and professors). The mean age was 23.26 (*SD* = 7.72) and the mean years of education was 14.27 (*SD* = 2.36). There was an overrepresentation of women in the sample (68/102, 66.67%). In terms of ethnicity, the majority (89/102, 87.26%) self-described as white or mestizo (mixed Amerindian and European ancestry), with the remainder identifying as being indigenous Americans. All were Spanish speakers.

Measures

The Word Accentuation Test (WAT) involves participants reading aloud 30 Spanish words in which there is an idiosyncratic stress pattern, but the accent marks indicating the pattern are not shown (Del Ser et al., 1997; Gomar et al., 2011). The accent marks in Spanish words indicate where the stress should be placed. Therefore, accurate pronunciation of words in which stress information is not given requires prior familiarity with the words, and hence recognized words would normally be read through the lexical reading route. Otherwise, words which are novel to the reader will be read from standard grammatical grapheme-phoneme conversion rules, which will usually produce errors. The WAT is considered to be a Spanish-language equivalent of English-language word pronunciation tasks, such as the National Adult Reading Test (Bright et al., 2018; Nelson & O'Connell, 1978) and the Test of Premorbid Function (Wechsler, 2011), which instead of stress marks rely on words with irregular spelling, such as 'paradigm'. Removal of stress marks is used in Spanish because irregularly pronounced words are very rare in the language. The words in the WAT are presented in isolation (i.e. no contextual information) in a list.

All 30 words of the WAT were displayed in upper-case on a single slide, in two columns on a tablet computer. The researcher awarded one point for each word pronounced correctly. Reliability data on the use of the WAT in Ecuador (where the current research was performed), revealed that the test has good internal consistency, Cronbach's α = .84, and test-retest reliability, r = .91 (Pluck, Almeida-Meza, Gonzalez-Lorza, Muñoz-Ycaza, & Trueba, 2017). Further reliability analyses focused on two alternatives to the WAT, the WAT–Sentences and the Stem Completion Implicit Reading Test, which are also reported below.

The first of these, the WAT-Sentences, is based on the idea that providing semantic contextual information could arguably improve performance as a measure of the lexical reading system. By placing each target word to be pronounced within a sentence, there should be coactivation of the words within the sentence which should aid retrieval by activating appropriate semantic lexicon entries. This is because word recognition is known to have cascade effects, in which neighboring words provide activation to enhance recognition (Boot & Pecher, 2008). There is in fact evidence that word pronunciation is enhanced by reading words within sentences, compared to reading in isolation. Beardsall (1998) took the 50 words from the English-language National Adult Reading Test and placed them in sentences. They found that poor to average readers improved significantly, compared to their ability pronouncing the same words shown in isolation. Interestingly, the effect was negligible for good readers, suggesting that it is not a general facilitation effect, but more likely a specific retrieval effect, most evident when word forms do actually exist within the individuals' lexical reading systems (otherwise they would still fail to pronounce correctly), but are not being retrieved efficiently.

To make the WAT-Sentences the same methodology as Beardsall (1998) was employed with adjustments to the Spanish-language context. The 30 words from the WAT were each placed within a sentence which provides semantic context. For example, one of the words in the WAT is ACOLITO, the correct pronunciation is with stress on the second vowel (ACÓLITO). In the WAT-Sentences the word was placed within a sentence: "*El montaguillo ayudaba al padre en la iglesia sirviendo como acolito en el altar*" (meaning "The altar boy was helping the priest in church as an acolyte at the altar"). The sentence is read aloud by the participant and the pronunciation of the target word only (in this case ACOLITO) is judged as correct or incorrect. The sentences were displayed on a tablet computer in a PowerPoint file with five on each slide. The participant read each of the 30 sentences aloud (one for each of the original 30 words in the WAT), and the researcher noted whether it was pronounced correctly or not. One point was awarded for each correct pronunciation.

Continuing with the concept of contextual facilitation, a third test was employed as a measure of naming ability. In this test, lexical access for target words is assessed, but for words in which there is no pronunciation ambiguity as in the WAT or WAT-Sentences tests. Furthermore, participants were provided with phonological and visual-semantic context, in addition to the orthographic-semantic context. To do this, 20 relatively low-frequency words were used as the targets. In a typical trial, the participant is presented with a photograph, and beneath that is a sentence in which the word representing a concept prominent in the photograph is incomplete. The first few letters are given, but the participant must complete the word. As an example, one trial has a photograph of a human humerus bone. The sentence below it is "*El profesor de anatomía describió los huesos como el hú*_______" (meaning "the anatomy professor described bones such as the hu_______". If the participant produces the correct word, i.e. "humerus" they receive one point. The slides, each comprising one photograph and one incomplete sentence, were shown one by one on a tablet computer in a PowerPoint file. The participants were required to read each sentence aloud and to complete the incomplete word. They were encouraged to guess any words that they were unsure of.

This test builds on the idea of measuring lexical contents while providing multimodal contextual information. Its creation was motivated by reports that patients with dementia can often recall words in memory tests at normal levels if they are completing the stems, i.e. the first letters are provided for them, despite abnormally low performance with recalling them without stems (Arroyo-Anllo, Beauchamps, Ingrand, Neau, & Gil, 2013). This effect is not confined only to implicit learning paradigms, but occurs additionally with standard naming tasks, in which patients with mild dementia do not differ from healthy controls if given semantic and phonemic cues (Balthazar, Cendes, & Damasceno, 2008). As this follows the literature on implicit memory, the assessment was labelled the Stem-Completion Implicit Reading Test (SCIRT).

Procedure

All participants were assessed individually in an interview room at the University. They all provided written informed consent, in accordance with the protocols approved by the local research ethics committee. The data collection for this study was performed in conjunction with two other studies. One a study of executive function tests and intelligence as predictors of academic achievement (Pluck et al., 2016) and the other a study establishing the validity and reliability of the WAT for use in Ecuador (Pluck et al., 2017). For all interviews, following the consent procedure the participants were interviewed about demographic factors such as age and ethnicity. They were also asked if they had ever received a formal diagnosis of dyslexia by a psychologist or other professional. Then they performed the WAT as described above, and then the SCIRT. There were then about 10 minutes of different cognitive tests not reported here, before the WAT-Sentences was performed. The WAT, SCIRT and WAT-Sentences tests were all displayed on a tablet computer, and recording of responses was by hand. A subset of 20 of the participants was assessed twice on the same tests, separated by four weeks. In each case the assessments were performed in the same sequence, in the same location and by the same researcher. This was to provide the test-retest estimates for the different measures. The subset of 20 participants who did the retest were mainly University employees (16/20), 12 males, with a men age of 34.30 (*SD* = 10.47) and had a mean of 13.6 (*SD* = 4.33) years of education.

After completion of the full assessments, the participants were thanked for their participation and debriefed. Sixty-four of the participants received course credit for volunteering for the research. The remainder were paid $20.

Approximately 12 months after participation in the research, the University records for the 86 undergraduate student participants were accessed. For each the total Grade Point Average (GPA) for all courses taken at the University up to that time point was recorded. In this university GPA ranges between zero and four with four being the best and highest possible score.

RESULTS

Reliability and validity of the language assessment tools

The psychometric properties of the WAT are already well established, and it is known to have good internal consistency, inter-rater reliability and stability, e.g. as revealed by test-retest reliability (Del Ser et al., 1997; Pluck et al., 2017). It is also known to be a valid measure of the lexical store, as shown by its very high correlation (r = .85) with the Vocabulary measure of the Wechsler Adult Intelligence Scale (Burin, Jorge, Arizaga, & Paulsen, 2000).

However, the two other tests, the WAT-Sentences and SCIRT, were developed specifically for this research and therefore their psychometric properties are unknown. Because of this, before examining the primary hypotheses concerning the correlations of the language tests with academic achievement, reliability and validity indices were computed.

The mean score on the SCIRT was 11.93 (*SD* = 2.98) and the median score was 12.00 (range 3-20). That the mean and median have similar values indicates minimal skewing of the data. For the WAT-Sentences similar results were found, although there was minor negative skewing with a mean score of 22.13 (*SD* = 4.00) and a median of 23.00 (range = 9-29). The internal consistencies of the two scales were assessed by Cronbach's alpha (α). For the WAT-Sentences the coefficient was α = 0.80. This would be considered to be 'good' internal consistency (Sharma, 2016). For the SCIRT the coefficient was somewhat lower at α = 0.71, nevertheless, this would still be considered as an 'acceptable' level of internal consistency. Secondly, test-retest reliability was examined. Such reliability is important because it reveals the stability of the measures. As it was assumed that these tests measure some aspects of the lexical route to speech, this should be stable if the test is capable of being a valid measure. The Pearson correlation is considered an appropriate method for such comparisons (Rousson, Gasser, & Seifert, 2002), however, rather than interpreting the actual r value, good test-retest reliability is established if the lower bound of the 95% confidence is at least 0.75. For both the WAT-Sentences and the SCIRT the test-retest reliabilities were high: WAT-Sentences $r = .92$, $p < .001$ (95% confidence intervals = 0.81 and 0.99) and SCIRT $r = .95$, $p < .001$ (95% confidence interval = 0.91 and 0.99). Both new tests therefore have good temporal stability.

To assess the concurrent validity of the two new measures (WAT-Sentences and SCIRT), their correlations with the WAT were examined, as this is accepted as a gold-standard measure of lexical entries. In the full sample, the correlation of the WAT with the WAT-Sentences was high, $r(102) = .86$, $p < .001$ (95% confidence intervals of $r = .79$ and .91). This strong correlation is perhaps not surprising considering that the target words to be pronounced in both tests are the same, the only difference being the reading context. The correlation of the SCIRT (quite a different test) with the WAT was somewhat lower, though still relatively high $r(102) =.52$, $p < .001$ (95% confidence intervals of r value = .35 and .67). Although somewhat smaller than the previous correlation, it has been argued that validity correlations are generally smaller than those for reliability as the concepts measured are emergent properties, and in such cases r values of greater than .5 can be interpreted as being "very high" correlations for the purpose of validation (Swank & Mullen, 2017). Therefore, it could be concluded that both the WAT-Sentences and the SCIRT are reliable instruments, and that they are valid measures of lexical ability.

Prediction of academic achievement

In the next step, it was examined whether the different language tests predicted academic achievement of the 86 university students, as measured by GPA. As this requires prediction, a linear regression was employed instead of correlation.

With the student GPA scores as the dependent variable and the WAT total scores as the independent variable, this produced a significant model predicting GPA, $F(1,84) = 16.11$, $p < .001$, with an R of 0.40. To predict GPA from the regression equation this would be GPA = 1.191 + .059*WAT scores. Next, the

WAT-Sentences task performance was analyzed in this same manner. This also produced a significant model predicting GPA scores, but the *R* value was a little lower than in the previous regression. For WAT-Sentences the model was $F(1,84) = 9.28$, $p = .002$, $R = .32$. By the regression equation GPA = 1.930 + .055*WAT-Sentences scores. As these analyses contain only one independent variable in each the *R* vales are equivalent to correlation *r* vales. To compare them directly it is perhaps easier to consider the R^2 values which show the shared variance between the reading tasks and GPA. For the WAT predicting GPA the R^2 is 0.161, indicating 16.1% shared variance between the variables. For the WAT-Sentences as a predictor of GPA the R^2 is 0.104, indicating only 10.4% of variance shared between the variables.

For the SCIRT, the stem-completion naming test, the independent variable this did not produce a significant model predicting GPA, $F(1,84) = .49$, $p = .487$ and the *R* value was only .076 ($R^2 = .006$). Therefore, there is very little shared variance between the variables and task performance cannot be used to predict academic achievement.

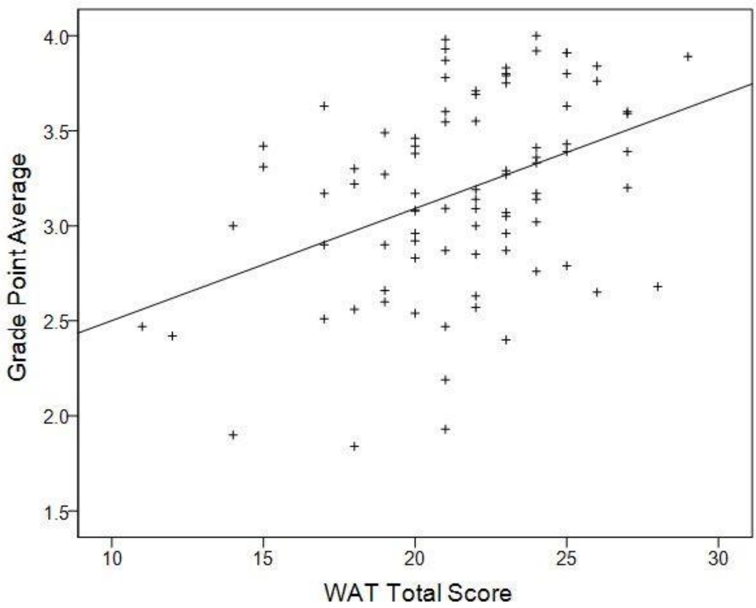

Figure 2. Scatterplot of the Word Accentuation Test (WAT) scores and Grade Point Average data for the Student Participants, including the regression line

As scores on both the WAT and WAT-Sentences were both significant predictors of GPA, in the next step the two were entered sequentially as independent variables in a multiple-linear regression analysis to predict GPA.

In this model with both variables entered, only WAT scores were significant predictors of GPA, *t* = 2.38, *p* = .020 with WAT-Sentences scores held constant. The WAT-Sentences scores with WAT scores held constant was *t* = .08, *p* = .933. The full model was significant, *F*(2,83) = 7.96, *p* = .001, with an *R* of .401, and an adjusted R^2 of .141. It is notable that *R* value with both reading tasks entered is exactly the same as with only the WAT scores entered (*R* = .401). Indeed, the change from WAT only scores to WAT and WAT-Sentences scores clearly did not significantly change predictive power of the model, with and R^2 change of 0.0, and a significance of the change at *p* = .933. This indicates that when the variance in WAT scores is counted, the WAT-Sentences explained no additional variance in GPA.

The significant relationships between word pronunciation performance and GPA are shown graphically in Figures 2 and 3, which include the regression equation lines. In the figures it can be observed that the range of test scores was reduced in the WAT-Sentences (range = 15-29) compared to the WAT (range = 11-29). This reduction in range probably explains some of the attenuation of the predictive value of GPA with WAT-Sentences scores (compared to the predictive values of WAT scores). It also notable that this difference in range between the WAT and WAT-Sentences tasks appears to affect the lower range of scores.

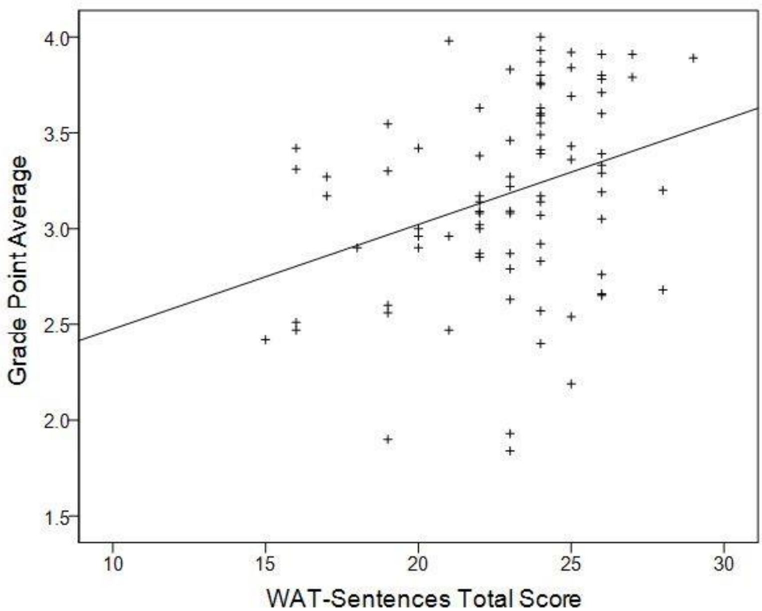

Figure 3. Scatterplot of the Word Accentuation Test (WAT) Sentences scores and Grade Point Average Data for the Student Participants, including the regression line

Comparison of performance between poor and good readers

This phenomenon described above appears to replicate that described by (Beardsall, 1998) in which lexical reading of words embedded in sentences improved pronunciation, but mainly in poor readers. This was explored further in the current data set. The median score on the WAT was 21.50 (range 9-29). By using a median-split on the whole sample, there were 51 low-scorers and 51 high-scorers. To assess facilitation by the semantic sentence context, a score was computed as the difference between reading scores for the WAT-sentences minus reading the normal WAT words. Thus, a high score would indicate improvement associated with reading the words within sentences. The low-scorers on the WAT did indeed improve significantly more than the high-scorers, (mean of 2.04, $SD = 2.44$ compared with a mean of 0.51, $SD = 1.61$), $t(100) = -.3.74$, $p < .001$, $d = .70$. This could of course be caused by there being more potential for improvement in people who started with a poor performance on the WAT test, compared to those who had good performance, and therefore, had fewer opportunities to improve. However, the significant effect remains if we calculate the dependent variable to be a ratio of the improvement to the number of errors made originally (i.e. the room for improvement), $t(100) = -2.40$, $p = .018$, $d = 0.46$.

As WAT low- and high-scorers (i.e. relatively poor and good readers) responded differently to the facilitating influence of semantic context, it is of interest to revisit the correlations of WAT-Sentences scores with GPA. As described above, the basic correlation of WAT-Sentences scores with GPA in the full sample was $R = .32$, $p = .002$. When we examine only the low-scorers the previously significant predictive model reduces to a non-significant model, $F(1,38) = .66$, $p = .422$, $R = .13$. For the high-scorers the model remained significant, $F(1,44) = 4.333$, $p = .043$, $R = .30$. Overall, this suggests that the semantic facilitation effect with low-scorers altered the relationship between task performance and GPA, effectively leaving no significant correlation. However, in those participants in which there was no facilitation by reading in sentences, the correlation between task performance and GPA is maintained.

The role of dyslexia

Finally, it was explored whether dyslexia diagnosis of the participants could explain the high association between WAT performance and GPA. Of the 102 participants, only 2 reported formal diagnoses of dyslexia. To examine if these participants performed poorly on the WAT we examined their *z* scores compared to the full sample. The WAT scores of neither participant would be considered abnormal; one had a *z* score of -0.682, and the other had a *z* score of 0.274. Clearly both where in the normal range of performances. Similarly, the *z* scores of their GPA scores were both within the normal range, being 0.228 and 1.028 respectively.

DISCUSSIONS

As hypothesized, the results suggest that lexical reading ability is a significant predictor of academic achievement at university level. The value of R = .401 for WAT scores predicting GPA may seem rather small. However, it should be taken within the context of other cognitive tests which tend to correlate very little with academic achievement in higher education. In fact, there are several studies reporting on intelligence tests and achievement in higher education which failed to find any association at all (e.g., Furnham, Chamorro-Premuzic, & McDougall, 2003). Some do find significant associations, but these tend to be considerably smaller than the association reported here. A systematic review and metanalysis of such studies estimated the mean correlation between intelligence test scores and university-level achievement was only .20, with a 95% confidence interval of .16 to .24 (Richardson et al., 2012). The current finding therefore represents an improvement on this. It should also be noted that cognitive test predictors of academic grades are inherently restricted due to the fact that multiple factors beyond those of the individual's ability contribute to a student's grades, such as the variability in grade allocation.

The other hypothesis in the current research was whether the breadth of individuals' semantic-conceptual abilities would be a factor in the prediction of academic success. This is because lexical reading usually involves access to meaning, although it is not necessary, the reading process can occur without it. This is shown in Figure 1 which displays a contemporary cognitive theory on the processes involved with reading and naming (Gvion & Friedmann, 2016). In the current task reading in the WAT, which was associated with academic grades, would have to follow the lexical route involving the Orthographic Input Lexicon and the Phonological Output Lexicon. This route has bidirectional connections with the Semantic Lexicon and as such if semantic priming information was available, such as in the WAT-Sentences task, this could potentially aid performance. However, there was no evidence that the semantic context enhanced the relationship between lexical reading performance and academic achievement.

Furthermore, in a task in which naming was used, rather than reading, there was no association with academic achievement at all. Again, referring to Figure 1, visual naming has to involve conceptual and semantic processing systems and yet this does not appear to predict academic achievement. Together these observations suggest that it may be the extent of stored orthographic or phonological forms within the lexical reading system that is important in the association with academic achievement, not the associated semantic lexicon information.

Related to this was the secondary observation that semantic context seemed to improve lexical reading performance in some people. It seemed that poor readers improved when performing the format in which there was contextual information (in the WAT-Sentences task), compared to their performance on the basic form of the task (i.e. reading words in isolation on the WAT). However, participants with good lexical reading ability did not show this facilitation. This implies that skilled readers, at least on the WAT task, can rely on their stored lexical knowledge to read aloud, but that poorer readers utilize more contextual semantic reading. This replicates, in Spanish, the similar finding of Beardsall (1998), who showed that reading irregular English-language words in sentences enhances performance for poor to average readers, but not for good readers. An educational implication of this may be that poorer readers in particular may have better comprehension from reading texts in context, as advocated in the basal approach to language learning (Jeynes & Littell, 2000).

In fact, this phenomenon appears to also support the thesis that the breadth of the semantic lexicon as an individual difference metric is not the important factor in academic achievement, but that the breadth of the reading lexical stores is. This is because for the poor readers, who did utilize semantic processing in their reading aloud, there was no significant relationship of reading performance with academic achievement. That relationship was only discernible in the good readers, who appeared not to be utilizing semantic context to improve their lexical reading ability.

Possible explanations for the correlation between grades and lexical reading

The results are of course simply observational, and the causation is not clear. One possibility though is that the ability to learn lexical forms reflects the efficiency of the declarative learning system. This suggestion is supported by studies with children, in which the development of the lexicon has been shown to correlate significantly with performance on tests on declarative learning ability (Kidd, 2012; Lum et al., 2012). A simple and quite reasonable alternative is that the breadth of a given individual's lexical stores reflects the quality of their education to date, those who have received the best education, and perhaps worked hardest, will tend to have both good lexical knowledge and good grades. Lexical reading may therefore be a particularly strong measure of the quality and depth of an individual's formal education.

Failure of the lexical reading route, when it is at a pathological level results in surface dyslexia, which is well-recognized as an acquired neuropsychological impairment, but also occurs in some cases of developmental dyslexia (Peterson et al., 2014). Therefore, one reason for the currently observed association between academic achievement and lexical reading ability could be that the WAT test is picking up sub-threshold dyslexia, which could then be linked to difficulties in the academic environment. Although the current research cannot fully address this, the results do not provide any support for the thesis, as of the two participants in the current sample who reported formal diagnoses of dyslexia, neither was performing abnormally in terms of academic achievement nor in WAT test performance.

Nevertheless, as the poor readers in the current study were essentially demonstrating some features of surface dyslexia, and associated with this was poor academic performance, then sub-threshold dyslexia remains a possible explanation for the current findings. Whatever the reason for the association between GPA and WAT test performance, it suggests that the breadth of lexical reading stores may have significant overlap with high-level educational achievement and may be worthy of further investigation which could elucidate the mechanisms.

Other interesting features of lexical reading as a cognitive ability

That lexical word reading is a predictor of academic achievement at university level adds to the corpus of observations indicating several curious and useful features of lexical reading as a cognitive skill. These have been researched mainly within the field of neuropsychology where several tests of lexical word reading are widely used for both clinical and research purposes, such as the National Adult Reading Test (Bright et al., 2018; Nelson & O'Connell, 1978) and the Test of Premorbid Function (Wechsler, 2011), and in several languages other than English, such as Spanish (Del Ser et al., 1997; Gomar et al., 2011), Portuguese (Alves et al., 2012) and Swedish (Rolstad et al., 2008). One feature is lexical reading’s remarkably high correlation with psychometrically measured intelligence, typically with correlations over .70 (Alves et al., 2012; Pluck et al., 2017; Schretlen, Buffington, Meyer, & Pearlson, 2005). Furthermore, the association with intelligence is stable over very long time periods, in one study over 66 years between intelligence assessment and reading assessment (Crawford, Deary, Starr, & Whalley, 2001). It also has high intrasubject stability of task performance in healthy participants, with test-retest correlations topically over .90 (Del Ser et al., 1997; Pluck et al., 2017). This stability even extends to the presence of neurological illness, in which the performance of lexical reading seems to be generally unaffected, unlike most other cognitive functions (Pluck et al., 2017; Rolstad et al., 2008). Such stability of the emergent property measured by the WAT could possibly explain its high association with academic achievement, compared to other cognitive tests. However, in the current research the psychometric properties of the alternative tests used here, the WAT-Sentences and the SCIRT, were deliberately explored. These were both found to have similar psychometric properties to the WAT. They differed principally on their involvement of semantic-lexical content, but not on stability, such as measured by test-retest reliability.

Limitations of the current findings

Nevertheless, there are some limitations to the current research which should be mentioned. One is the generalizability. The current research was conducted in Spanish, with words with unusual stress patterns used to force reading to be via the lexical route. Most of the research on this topic has been conducted in the English language, in which irregular-spelt words are employed. Although reading with correct stress patterns is often considered equivalent to reading irregular-spelt words, there may be some differences. It would be of interest to see if the current results can be replicated in a language such as English in which irregular matches between orthography and phonemes are common.

Another limitation is related to the finding that academic achievement can be predicted from lexical reading, and that this relationship is stronger that typically found in higher education samples. However, there is no direct comparison with other types of test, such as fluid intelligence or executive functioning. It would be of particular interest to see whether lexical reading really is a better predictor compared to other tests, as a possible interpretation is that lexical reading is simply a strong measure of the *g* factor (i.e., general intelligence). Further studies could include non-linguistic tasks, particular those with high *g* loadings. Furthermore, as the current findings are purely observational and correlational, there could be other causal agents involved which are driving the associations reported here. Although there is a significant correlation between lexical reading and GPA, no causality should be inferred. Nevertheless, the current results provide a novel observation and provide stimulus for future studies that can explore the phenomenon in more detail.

Conclusions. Reading aloud of words that require lexical entries to allow correct pronunciation is a significant predictor of high-level academic achievement. And this appears not to be driven by individuals with good lexical reading also having better semantic lexical stores (e.g. conceptual knowledge). The reason that lexical reading predicts academic performance remains unclear but viable hypotheses are that lexical reading is a sensitive way to detect subthreshold dyslexia, that it indicates the efficiency of declarative memory systems, or perhaps it simply provides a good measure of the quality and quantity of past formal education. Such lexical processes and their study as predictors of real-life performance present an interesting language-system based alternative to attempts to predict performance from cognitive domain-general processing systems (e.g. fluid intelligence or executive functioning).

REFERENCES

Alves, L., Simoes, M. R., & Martins, C. (2012). The estimation of premorbid intelligence levels among Portuguese speakers: the Irregular Word Reading Test (TeLPI). *Archives of Clinical Neuropsychology, 27*(1), 58-68.

Arroyo-Anllo, E. M., Beauchamps, M., Ingrand, P., Neau, J. P., & Gil, R. (2013). Lexical priming in Alzheimer's disease and aphasia. *European Neurology*, *69*(6), 360-365.

Baddeley, A. (2003). Working memory and language: an overview. *Journal of Communication Disorders, 36*(3), 189-208.

Baddeley, A. (2017). *Exploring working memory: Selected works of Alan Baddeley*. London, UK: Routledge.

Baddeley, A. D., & Hitch, G. (1974). Working Memory. In G. H. Bower (Ed.), *Psychology of Learning and Motivation: Advances in Research and Theory*. New York: Academic Press.

Baldo, J. V., Bunge, S. A., Wilson, S. M., & Dronkers, N. F. (2010). Is relational reasoning dependent on language? A voxel-based lesion symptom mapping study. *Brain and Language, 113*(2), 59-64.

Balthazar, M. L., Cendes, F., & Damasceno, B. P. (2008). Semantic error patterns on the Boston Naming Test in normal aging, amnestic mild cognitive impairment, and mild Alzheimer's disease: is there semantic disruption? *Neuropsychology, 22*(6), 703-709.

Barbey, A. K., Colom, R., Solomon, J., Krueger, F., Forbes, C., & Grafman, J. (2012). An integrative architecture for general intelligence and executive function revealed by lesion mapping. *Brain, 135*(4), 1154-1164.

Beardsall, L. (1998). Development of the Cambridge Contextual Reading Test for improving the estimation of premorbid verbal intelligence in older persons with dementia. *British Journal of Clinical Psychology, 37*(2), 229-240.

Binet, A. (1903). *L'Etude experimentale de l'intelligence*. Paris: Schleicher Freres & Cie.

Blank, I., & Fedorenko, E. (2017). Domain-general brain regions do not track linguistic input as closely as language-selective regions. *Journal of Neuroscience, 37*, 9999-10011.

Boot, I., & Pecher, D. (2008). Word recognition is affected by the meaning of orthographic neighbours: Evidence from semantic decision tasks. *Language and Cognitive Processes, 23*(3), 375-393.

Bright, P., Hale, E., Gooch, V. J., Myhill, T., & van der Linde, I. (2018). The National Adult Reading Test: restandardisation against the Wechsler Adult Intelligence Scale-Fourth Edition. *Neuropsychological Rehabilitation, 28*(6), 1019-1027.

Burin, D. I., Jorge, R. E., Arizaga, R. A., & Paulsen, J. S. (2000). Estimation of premorbid intelligence: the Word Accentuation Test- Buenos Aires version. *Journal of Clinical and Experimental Neuropsychology, 22*(5), 677-685.

Campbell, K. L., & Tyler, L. K. (2018). Language-related domain-specific and domain-general systems in the human brain. *Current Opinion in Behavioral Sciences, 21*, 132-137.

Crawford, J. R., Deary, I. J., Starr, J., & Whalley, L. J. (2001). The NART as an index of prior intellectual functioning: a retrospective validity study covering a 66-year interval. *Psychological Medicine, 31*, 451-458.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior, 19*(4), 450-466.

Del Ser, T., Gonzalez-Montalvo, J. I., Martinez-Espinosa, S., Delgado-Villapalos, C., & Bermejo, F. (1997). Estimation of premorbid intelligence in Spanish people with the Word Accentuation Test and its application to the diagnosis of dementia. *Brain and Cognition, 33*(3), 343-356.

Formoso, J., Injoque-Ricle, I., Barreyro, J. P., Calero, A., Jacubovich, S., & Burín, D. I. (2018). Mathematical cognition, working memory, and processing speed in children. *Cognition, Brain, Behavior. An Interdisciplinary Journal, 22*(2), 59-84.

Furnham, A., Chamorro-Premuzic, T., & McDougall, F. (2003). Personality, cognitive ability, and beliefs about intelligence as predictors of academic performance. *Learning and Individual Differences, 14*, 49-66.

Gomar, J. J., Ortiz-Gil, J., McKenna, P. J., Salvador, R., Sans-Sansa, B., Sarro, S., . . . Pomarol-Clotet, E. (2011). Validation of the Word Accentuation Test (TAP) as a means of estimating premorbid IQ in Spanish speakers. *Schizophrenia Research, 128*(1-3), 175-176.

Gvion, A., & Friedmann, N. (2016). A principled relation between reading and naming in acquired and developmental anomia: Surface dyslexia following impairment in the phonological output lexicon. *Frontiers in Psychology, 7*.

Higgins, D. M., Peterson, J. B., Pihl, R. O., & Lee, A. G. (2007). Prefrontal cognitive ability, intelligence, Big Five personality, and the prediction of advanced academic and workplace performance. *Journal of Personality and Social Psychology, 93*(2), 298-319.

Iglesias-Sarmiento, V., López, N. C., & Rodríguez, J. L. R. (2015). Updating executive function and performance in reading comprehension and problem solving. *Anales de Psicología, 31*(1), 298-309.

Jeynes, W. H., & Littell, S. W. (2000). A meta-analysis of studies examining the effect of whole language instruction on the literacy of low-SES students. *The Elementary School Journal, 10*(1), 21-33.

Kaufman, S. B., DeYoung, C. G., Gray, J. R., Jiménez, L., Brown, J., & Mackintosh, N. (2010). Implicit learning as an ability. *Cognition, 116*(3), 321-340.

Kidd, E. (2012). Implicit statistical learning is directly associated with the acquisition of syntax. *Developmental Psychology, 48*(1), 171-184.

Lewis, J. W. (2006). Cortical networks related to human use of tools. *Neuroscientist, 12*(3), 211-231.

Lum, J. A., Conti-Ramsden, G., Page, D., & Ullman, M. T. (2012). Working, declarative and procedural memory in specific language impairment. *Cortex, 48*(9), 1138-1154.

Mardh, S., Nagga, K., & Samuelsson, S. (2013). A longitudinal study of semantic memory impairment in patients with Alzheimer's disease. *Cortex, 49*(2), 528-533.

Nelson, H. E., & O'Connell, A. (1978). Dementia: The estimation of premorbid intelligence levels using the New Adult Reading Test. *Cortex, 14*(2), 234-244.

Peterson, R. L., Pennington, B. F., Olson, R. K., & Wadsworth, S. J. (2014). Longitudinal stability of phonological and surface subtypes of developmental dyslexia. *Scientific Studies of Reading, 18*(5), 347-362.

Pinel, P., & Dehaene, S. (2010). Beyond hemispheric dominance: Brain regions underlying the joint lateralization of language and arithmetic to the left hemisphere. *Journal of Cognitive Neuroscience, 22*(1), 48-66.

Pluck, G., Almeida-Meza, P., Gonzalez-Lorza, A., Muñoz-Ycaza, R. A., & Trueba, A. F. (2017). Estimación de la función cognitiva premórbida con el Test de Acentuación de Palabras. *Revista Ecuatoriana de Neurología, 26*(3), 226-234.

Pluck, G., Bravo Mancero, P., Maldonado Gavilanez, C. E., Urquizo Alcívar, A. M., Ortíz Encalada, P. A., Tello Carrasco, E., . . . Trueba, A. F. (in press). Modulation of striatum based non-declarative and medial temporal lobe based declarative memory predicts academic achievement at university level. *Trends in Neuroscience and Education*.

Pluck, G., Ruales-Chieruzzi, C. B., Paucar-Guerra, E. J., Andrade-Guimaraes, M. V., & Trueba, A. F. (2016). Separate contributions of general intelligence and right prefrontal neurocognitive functions to academic achievement at university level. *Trends in Neuroscience and Education, 5*(4), 178-185.

Premack, D. (2007). Human and animal cognition: continuity and discontinuity. *Proceedings of the National Academy of Sciences of the United States of America, 104*(35), 13861-13867.

Rasmussen, A., & Lingard, B. (2018). Excellence in education policies: Catering to the needs of gifted and talented or those of self-interest? *European Educational Research Journal*, 1474904118771466.

Richardson, M., Abraham, C., & Bond, R. (2012). Psychological correlates of university students' academic performance: a systematic review and meta-analysis. *Psychological Bulletin*, *138*(2), 353-387.

Rolstad, S., Nordlund, A., Gustavsson, M. H., Eckerstrom, C., Klang, O., Hansen, S., & Wallin, A. (2008). The Swedish National Adult Reading Test (NART-SWE): a test of premorbid IQ. *Scandinavian Journal of Psychology, 49*(6), 577-582.

Rousson, V., Gasser, T., & Seifert, B. (2002). Assessing intrarater, interrater and test-retest reliability of continuous measurements. *Statistics in Medicine, 21*(22), 3431-3446.

Santarnecchi, E., Emmendorfer, A., & Pascual-Leone, A. (2017). Dissecting the parieto-frontal correlates of fluid intelligence: A comprehensive ALE meta-analysis study. *Intelligence, 63*, 9-28.

Schretlen, D. J., Buffington, A. L., Meyer, S. M., & Pearlson, G. D. (2005). The use of word-reading to estimate "premorbid" ability in cognitive domains other than intelligence. *Journal of the International Neuropsychological Society, 11*(6), 784-787.

Sharma, B. (2016). A focus on reliability in developmental research through Cronbach's Alpha among medical, dental and paramedical professionals. *Asian Pacific Journal of Health Sciences, 3*(4), 271-278.

Squire, L. R., & Dede, A. J. (2015). Conscious and unconscious memory systems. *Cold Spring Harbor Perspectives in Biology, 7*(3), 1-14.

St Clair-Thompson, H. L., & Gathercole, S. E. (2006). Executive functions and achievements in school: shifting, updating, inhibition, and working memory. *Quarterly Journal of Experimental Psychology, 59*(4), 745-759.

Storms, G., Saerens, J., & De Deyn, P. P. (2004). Normative data for the Boston Naming Test in native Dutch-speaking Belgian children and the relation with intelligence. *Brain and Language*, *91*(3), 274-281.

Stout, D., & Chaminade, T. (2007). The evolutionary neuroscience of tool making. *Neuropsychologia*, *45*(5), 1091-1100.

Swank, J. M., & Mullen, P. R. (2017). Evaluating evidence for conceptually related constructs using bivariate correlations. *Measurement and Evaluation in Counseling and Development, 50*(4), 270-274.

Ullman, M. T. (2013). The role of declarative and procedural memory in disorders of language. *Linguistic Variation, 13*(2), 133-154.

Vingerhoets, G., Alderweireldt, A. S., Vandemaele, P., Cai, Q., Van der Haegen, L., Brysbaert, M., & Achten, E. (2013). Praxis and language are linked: evidence from co-lateralization in individuals with atypical language dominance. *Cortex*, *49*(1), 172-183.

Wechsler, D. (2011). *Test of Premorbid Functioning. UK version (TOPF UK)*. Oxford: UK: Pearson Corporation.

Woolgar, A., Duncan, J., Manes, F., & Fedorenko, E. (2018). Fluid intelligence is supported by the multiple-demand system not the language system. *Nature Human Behaviour*, *2*(3), 200-204.