Pluck, G. & Ruales-Chieruzzi, C. B. (in press). Estimation of premorbid intelligence and executive cognitive functions with lexical reading tasks. *Psychology & Neuroscience*.

Estimation of Premorbid Intelligence and Executive Cognitive Functions with Lexical Reading Tasks

Graham Pluck[1,2] and Carlos B. Ruales-Chieruzzi[1]

[1] Institute of Neurosciences, Universidad San Francisco de Quito

[2] Faculty of Psychology, Chulalongkorn University

Author Note

Graham Pluck https://orcid.org/0000-0002-0368-0051

Twitter: @DrGPluck

The authors report no conflict of interest. We would like to thank those who contributed to data collection: Dr Patricia Bravo in Riobamba, Dr Amalín Mayorga Albán in Guayaquil, Jose Hernandez Rodriguez in Manta, and Sarahí Pontón and Pamela Almeida Meza in Quito. This research was funded by a grant from the Universidad San Francisco de Quito Chancellor's Grant Scheme. WAT and MMSE test data reported in Study 1, from a subsample of the participants, was previously analyzed in Pluck et al. (2018). The data that support the findings of this study are openly available in PsychArchives at http://dx.doi.org/10.23668/psycharchives.2897

Correspondence concerning this article should be addressed to Dr Graham Pluck, School of Social Sciences and Humanities, Universidad San Francisco de Quito, Diego de Robles y Vía Interoceánica, P.O.BOX: 170901. Tel: +,593 2 297-1700, fax: Fax: +593 2 289-0070, email: g.c.pluck@gmail.com

Abstract

***Objective*:** Estimation of premorbid function is essential to accurate identification of cognitive impairments. We explored how lexical tasks can be used to estimate various cognitive functions.

***Methods*:** Adults with cognitive impairment due to neurological illness (n = 15) and a non-clinical sample of adults (n = 143) completed tests of word pronunciation, lexical decision, and stem-completion naming. In three studies lexical-task performance associations with intelligence (WAIS-IV), episodic memory, executive functioning, and theory of mind were explored.

***Results*:** At the group level, word pronunciation was robust in the presence of cognitive impairment. However, as a case series, patients showed idiosyncratic patterns of preservation of lexical skills. All of the lexical tasks were highly correlated with IQ scores in the non-clinical sample, suggesting that they may function well as estimators of premorbid intelligence. Simulated impairments in non-clinical adults revealed that the median score from the three different tasks had the highest correlation with, and provided the most accurate and precise estimates of, measured IQ, and was also the least sensitive to impairment. We also show that these methods predict executive functions, in particular, proverb interpretation, phonemic/semantic alternating verbal fluency, and working memory span performance.

***Conclusions*:** Several lexical tasks are potentially useful in predication of pre-illness cognitive ability. However, due to the heterogeneity of impairments between patients, estimation of premorbid levels could be improved by the use of average estimated values from multiple tests. This could potentially improve diagnostic accuracy and quantification of neuropsychological impairments.

Keywords: premorbid function, intelligence, executive function, reading, mental lexicon

Public Significance Statement

Understanding a patient's cognitive impairments following brain illness is better if their abilities before the illness can be estimated.

Reading that requires preexisting knowledge of the words, such as how to pronounce them, can be used for this purpose.

Word pronunciation, particularly when combined with other tests of word knowledge, appears to be not much affected by cognitive impairment, and gives a good indication of a patient's pre-illness ability.

Estimation of Premorbid Intelligence and Executive Cognitive Functions with Lexical Reading Tasks

Accurate estimation of premorbid cognitive ability is very important in clinical practice and research in clinical neurosciences, particularly neuropsychology. The comparison of premorbid estimations of function to observed levels in the pathological state allows clinicians to evaluate the extent of impairment at the individual level, allowing personalized quantification and definition of impairments. Several methods of premorbid estimation of cognitive function already exist, however, any advances or improvements in the use of existing tools would have widespread benefits.

There is substantial variation in human cognitive abilities, to the extent that between individuals, levels of cognitive performance in pathological states frequently overlap with performance in health. Clearly, personalized estimation of premorbid level is better than comparison to the estimated national norm for healthy people. Such methods also have widespread uses in clinical neuroscience research. The most common application is to match control and patient groups for premorbid level, such that any differences in current performance can be inferred to have been caused by the illness. However, they can also be used to address other research issues such as ‘which came first’ questions, for example, cognitive profiles associated in people with substance abuse histories (e.g., Pluck et al., 2012). A further application, is to measure cognitive reserve, an important predictor of dementia prognosis (Sobral et al., 2015).

Methods of premorbid estimation that have been employed include clinical opinion of premorbid level, or more formally calculated estimations based on demographics (Apolinario et al., 2013; Crawford et al., 2001) or the use of tests from common cognitive batteries that tend to ‘hold’ (de Oliveira et al., 2014), or the best performance across multiple subtests

(Vanderploeg et al., 1996). However, these have all tended to provide estimates that suffer from biases, either underestimating or overestimating performance in patients with known cognitive impairments, such as in dementia.

Probably the most successful attempts to estimate cognitive performance have been with the use of lexical tasks. Vocabulary is highly associated with other cognitive functions, in fact, the Vocabulary assessment in Wechsler intelligence tests is the best single measure of general intelligence, known as the *g* factor (Jensen, 2001). In addition, a task of pronouncing low-frequency irregularly spelled words, the National Adult Reading Test (NART; Nelson & Willison, 1991), has an equivalently high load on the *g* factor (Crawford et al., 1989). Accordingly, word pronunciation tasks have consistently been found to correlate very highly with standardized intelligence tests (Del Ser et al., 1997; Gil et al., 2019; Matsuoka et al., 2006; Nelson & Willison, 1991; Wechsler, 2001).

Neuropsychological models of word pronunciation distinguish multiple routes from the visually presented word to speech output, not all of which access the mental lexicon (Pluck, 2018). Forcing people to read through a lexical route allows an estimation of the lexicon size, and it is this aspect that seems to correlate so highly with intelligence. Cognitive tests that require access to the mental lexicon include, confrontation naming, lexical decision tasks, and single word pronunciation tasks (for material that cannot be accurately decoded by grapheme-phoneme conversion rules). In English, the latter method is usually achieved with the pronunciation of irregularly spelled words (Nelson & Willison, 1991; Wechsler, 2001, 2011). However, the principle works well in languages which lack that feature. In languages that have few irregularly pronounced words, such as Spanish or Portuguese, lexically-driven pronunciation can be forced by using words which require a particular stress, but for which the stress pattern information (i.e., the accent marks) has been removed (e.g., Del Ser et al., 1997; Gil et al., 2019). Or by using loan words, which do have irregular pronunciation, such as in the Swedish version of the NART (Rolstad et al., 2008).Compound kanji, for which the pronunciation specific to each kanji combination must be known for successful pronunciation, is used in Japanese (Matsuoka et al., 2006). Thus, lexical single-word reading is widely used to estimate intelligence test scores around the world. In theory, as lexical reading is highly associated with the *g* factor, it could be possible to predict scores on tests other than intelligence. In particular, executive control is highly related to intelligence, and it has even been said that the latent variable of executive function is identical to the *g* factor (Royall & Palmer, 2014). Nevertheless, this aspect of premorbid estimation of cognitive function has rarely been explored.

To function as premorbid estimators, tests must also be robust in the presence of impaired states of brain function, that is, they tend to ‘hold’. In fact, single word pronunciation tasks achieve this second criteria well, for example, being little affected by traumatic brain injury (Green et al., 2008). And despite losses to language skills in dementias such as Alzheimer’s disease (Martínez-Nicolás et al., 2019), pronunciation information seems to be preserved (Del Ser et al., 1997; Pluck et al., 2018; Rolstad et al., 2008). This may be because although pronunciation of irregular words is predicated on the word having existed in the declarative mental lexicon, the actual articulation of irregular words seems to be inherently procedural. Commonly used single-word pronunciation tests of premorbid IQ include the Wechsler Test of Adult Reading (Wechsler, 2001), Test of Premorbid Functioning (Wechsler, 2011), and National Adult Reading Test (Nelson & Willison, 1991) in English, and the Word Accentuation Test (WAT; Del Ser et al., 1997) in Spanish.

Considering their importance in clinical neurosciences, they are very simple tests, taking only three or four minutes to administer. Furthermore, there are some fallibilities to single word pronunciation tasks. Pronunciation does not fully hold: there is some drop in performance associated with dementias, such as Alzheimer’s disease (Stott et al., 2017), and with traumatic brain injury (Morris et al., 2005). Furthermore, the extent to which different neurodegenerative disorders affect word pronunciation seems to vary, with Korsakoff's syndrome appearing to be one of the most sensitive to impairment (Heirene et al., 2018). As a specific cognitive skill, reading words aloud is also sensitive to focal neural insults that impair those functions. Finally, single word pronunciation is limited in that it cannot be used in patients with speech disorders, such as dysarthria, which affects about 42% of all stroke patients (Lawrence et al., 2001).

There have been various attempts to improve on the single word pronunciation tasks. One approach is to minimize the effects of neurological illness on performance. Although language is clearly affected in dementia, such as that associated with Alzheimer's disease (Martínez-Nicolás et al., 2019), at least in the earlier stages, language problems may be due to retrieval difficulties, rather than damage to the semantic system (Balthazar et al., 2008; Nebes et al., 1984). As retrieval of single words is known to be enhanced by semantic activation of neighboring words (Boot & Pecher, 2008), it should be possible to provide additional activation to the underlying representations of the stimuli words, providing they are already in the lexicon. This could potentially minimize reading errors caused by neurological illness. This has been attempted in the Cambridge Contextual Reading Test (Beardsall & Huppert, 1994), which used the words from the NART in sentences. This does indeed produce better performance in patients with Alzheimer's disease suggesting that it may be a better measure of premorbid ability than single word pronunciation (Conway & O'Carroll, 1997). On the same theme of maximizing patient performance, several studies have shown that patients with dementia can show unimpaired levels of word recall when they are tested with stem-completion rather than having to recall the whole word (Arroyo-Anlló et al., 2013). Furthermore, this effect is not limited to memory tests, even mild-stage Alzheimer's disease patients show unimpaired performance on the Boston Naming Test if they are given semantic and phonemic cues (Balthazar et al., 2008). Thus, stem-completion with cues could potentially be used as a hold test. To our knowledge, this has not been previously explored.

Another option is to use lexical decision instead of word pronunciation, such as asking participants to select the real word from a set in which only one is real. As this still measures the mental lexicon, it could be resistant to neurological illness even if there is semantic knowledge loss, as recognition of the wordform is all that is needed, not recall of the meaning. It also allows for testing of patients with speech disorders. Several tests of this nature have been developed, such as the English-language Spot-the-Word Test (Baddeley et al., 1992) and the Spanish Lexical Decision Task (Pluck, 2020).

Two studies have directly compared the NART (single word pronunciation), Cambridge Contextual Reading Test (pronunciation in sentences) and the Spot-the-Word Test (lexical decision) and how well they hold in the presence of dementia. One reported that the Spot-the-Word Test was significantly worse in a group of patients with mild to moderate dementia (no differential diagnoses supplied) compared to controls, but that both of the pronunciation tasks were performed normally (Beardsall & Huppert, 1997). However a later study using similar methods reported the opposite, that both pronunciation tasks appeared to be affected by severity of cognitive impairment in Alzheimer’s’ disease, but the lexical decision task was not (McFarlane et al., 2006). It seems likely that variation in cognitive symptoms between the two study samples is responsible for the discrepancy.

In the current research we explored these methods as potential premorbid measures of cognitive function: single word pronunciation, pronunciation in sentences, and lexical decision. However, we also investigated the potential of stem-completion naming as a premorbid measure. Given the equivocal findings on preservation of function in the presence of cognitive impairment, we also explored whether variation in how well the different tests hold may vary between patients with cognitive impairments, and if so, whether this could be exploited to improve the accuracy and precision of estimates of premorbid functioning. Finally, we performed an exploratory study into how well cognitive test scores other than IQ can be predicted, focusing on tests of executive functions.

Study 1: Resistance to Cognitive Impairment of Lexical Tasks

Method

Design and Participants

The aim of this study was to examine the effects of cognitive impairments caused by neurological illness on different cognitive tests. In particular, which lexical tests are least and most affected by cognitive impairment associated with neurological illness. This used standard patient-control group comparisons. In addition, we aimed to examine performance on a case-by-case basis in the neurological patients to assess how different potential hold tests are preserved in the presence of cognitive impairment.

To this end we studied 15 patients with cognitive impairment, 6 males, mean age = 81.47 (SD = 5.72, range = 77 – 88), and 12 healthy control participants, 4 males, mean age 82.58 (SD = 6.57, range = 70 – 93). The years of completed formal education was equivalent between groups, patient mean = 12.71 (SD = 3.54), control mean = 13.17 (SD = 3.16). The cognitive impairment group had a mean Mini-mental State Examination score of 19.13 (SD= 4.75) and the elderly control group had a mean of 26.83 (SD = 1.90), a significant difference, $t(19.180) = 5.73$, $p < .001$, $D = 5.06$.

Fourteen of the cognitive impairment group were recruited and tested at a private center providing day care for people with dementia, located in Quito, Ecuador. Formal diagnoses were generally not made by the clinically-involved neurologists or psychiatrists. The clinical neuropsychologist at the day center where the patients were recruited provided several diagnoses. She suggested nine had probable Alzheimer’s disease (AD), one had probable vascular dementia (VaD), one had probable frontotemporal dementia (FTD), and one had mild cognitive impairment (MCI). As the diagnoses were not provided by the responsible clinician, we have described these as ‘probable’ diagnoses. Two other patients did not receive any diagnosis other than 'senile dementia' which was recorded in their medical notes. One individual who was recruited as a control participant revealed a diagnosis of Parkinson's disease during the interview. As even early stage Parkinson's disease is associated with cognitive impairment (Elgh et al., 2009), this participant was analyzed in the cognitive impairment group. The available diagnoses are shown in Table 1. The control group was recruited in the same geographical area as the day care center and were family members and acquaintances of the research team or were living a in residential care home. In both groups, recruits were from high socioeconomic status backgrounds. The members of the control group reported no neurological illnesses. All were Spanish speakers.

[Table 1 near here]

Assessments

Potential 'Hold' Tests. Four lexical reading tests were employed as potential estimators of premorbid function. Three of the tests are methods that have previously been explored as estimators of premorbid function, and one is a new task, the Stem Completion Implicit Reading Test.

Word Accentuation Test (WAT). This is the most widely-used Spanish language test used to estimate premorbid cognitive ability (Del Ser et al., 1997; Sierra Sanjurjo et al., 2015). The WAT (also known in Spanish as the *Test de Acentuación de Palabras*) involves participants reading and pronouncing a list of 30 words. It is equivalent to the English-language NART (Nelson & Willison, 1991). The WAT is scored as one point for each word pronounced correctly. The WAT has been shown to be resistant to dementia (Del Ser et al., 1997), and to have good psychometric properties including internal consistency and retest reliability in the Ecuadorian context (Pluck et al., 2018).

Word Accentuation Test- Sentences (WAT-Sentences). This is an experimental task in which the words from the WAT are embedded in sentences. It is equivalent to the Cambridge Contextual Reading Test (Beardsall & Huppert, 1994). We produced one sentence for each target word from the WAT. These sentences were presented on pages in a booklet with five sentences on each page. Scoring was exactly the same as for the WAT, one point for each target word pronounced correctly. We have previously reported acceptable internal consistency and retest reliability for the WAT-Sentences in an Ecuadorian sample (Pluck, 2018).

Spanish Lexical Decision Task (SpanLex). This involves participants selecting the single real word from sets of three words. The other two in each set are strings of letters that look like legitimate words but do not exist in the language. It is a Spanish equivalent to the English-language Spot-the-Word-Test (Baddeley et al., 1992). In the full SpanLex there are 36 word-triplets. The test has good internal consistency and retest reliability, established in Ecuador (Pluck, 2020). We used an abbreviated version with 17 word-triplets which were approximately every odd numbered item from the full test. This abbreviated version was used to reduce the testing load on the patients with dementia. The controls completed the full-version but only the 17 items from the abbreviated version are used in between group comparisons. The SpanLex is available to download from http://www.gpluck.co.uk.

Stem Completion Implicit Reading Test (SCIRT). This is an experimental measure, designed to assess contents of the mental lexicon by using a stem completion approach, with semantic, visual, and phonemic cues. Participants are shown a photograph and a sentence in which the final word is incomplete. Their task is to complete the sentence. For example, a photo of coral, and asked to complete a sentence such as '*The reef fish swim through the co…*'. There were 20 of these trials, all using relatively low frequency words. Each trial received one point if a word was produced that matched the photo and completed the sentence sensibly. We have previously shown that the SCIRT has adequate internal consistency and retest reliability (Pluck, 2018).

Tests of Current Cognitive Function. Two tests were employed to measure current cognitive performance, one of episodic memory and the other of fluid intelligence.

Test of Episodic Visual Incidental Learning (TEVIL). This is used to measure episodic memory (Pluck et al., 2019). It uses an incidental learning paradigm to focus on memory processes rather than deliberate learning strategies. There were 18 learning trials in which abstract designs were presented in a perceptual task. Later, after approximately 25 minutes, a surprise recognition test was given. In each recognition trial, one of the original 18 stimuli were shown, with two completely new abstract designs. The task was to identify which shape had been seen before. One point was awarded for each ‘old’ stimulus correctly identified, thus a maximum of 18 points were available.

Matrix Matching Test (MMT). As a measure of fluid intelligence we used the visuospatial tasks from the Matrix Matching Test, which has been shown to be valid and reliable in the Ecuadorian context (Pluck, 2019). This is similar to other visuospatial intelligence tests such as the Matrices subtest used in some of the Wechsler intelligence tests (e.g., Wechsler, 1999). There were 16 trials, in each a visual pattern was presented with a piece missing. The task was to choose which of six options completes the pattern. The tasks were graded in difficulty, from simple to complex. One point was awarded for each trial in which the correct response was selected. The task was stopped after three consecutive scores of zero.

Procedure

All participants provided informed written consent, in accordance with the ethics committee approved protocol. All participants performed all of the tests described above, with the incidental learning aspect of the TEVIL performed first, and the recall aspect performed last. Participants were debriefed and thanked for participation but did not receive any renumeration.

Result and Discussion of Study 1

The data supporting this study is publicly available at http://dx.doi.org/10.23668/psycharchives.2897. Mean scores were calculated for all of the cognitive measures and are shown graphically in Figure 1. All scores were normally distributed, based on skew and kurtosis (Kim, 2013), except WAT scores. One-tailed *t*-tests and a Mann-Whitney U-test (for WAT scores) were used to compare the cognitive impairment and elderly control groups, with effect size given as Glass' Delta (*D*), that is, the between group mean difference divided by the control group standard deviation. There was no significant difference between the cognitively impaired and elderly control groups for performance on the WAT, $U = 89.00$, $p = .491$, $D = .02$. However, for all of the other potential hold tests, the cognitively impaired group scored significantly below the elderly control group, with effect sizes of WAT-Sentences $D = 1.19$, SpanLex $D = 1.06$, and SCIRT $D = 1.60$. It is notable from Figure 1 that the hope that semantic context would aid pronunciation of the WAT stimuli words only seemed to apply to the control group, having no observable effect on the performance of the cognitively impaired group.

[Figure 1 near here]

Regarding the tests of current cognitive function, as expected, in both comparisons the cognitive impaired group scored significantly below the level of the elderly control group (both $p < .003$). The effect sizes were large, TEVIL episode memory $D = 2.53$, MMT fluid intelligence $D = 1.16$.

These results confirm previous observations that tests of single-word pronunciation seem to hold well in the presence of cognitive impairment. At the between group level, all of the other measures, be them potential estimators of premorbid function or current function, were significantly affected by the neurological illnesses in the cognitive impairment group. However, group comparisons such as this will miss individual cases in which word pronunciation is actually lower than would be expected. To examine the consistency of word pronunciation as a preserved ability, we examined the 15 participants with cognitive impairments as a case series. For each we calculated how far their individual scores on the different lexical tasks differed from the mean of the control group, measured in control group standard deviations. These figures are shown in Table 1, in which the best performance of each participant with cognitive impairment, relative to the control mean, is shown in bold. As expected, the WAT was the best overall, as 9/15 cognitive impaired patients had their best performance on that test. However, 3/15 had their best performance on the stem-completion test (SCIRT), 2/15 on the lexical decision task (SpanLex) and 1/15 on the WAT-Sentences test. It is also notable that 3/15 patients had their worst performance on the WAT.

The results from the case-series analysis suggest some limited potential for each of the three alternative methods of estimating premorbid function: WAT-Sentences, SpanLex and SCIRT. However, for the WAT-Sentences it may be problematic that semantic context tends to enhance performance in healthy controls, but not in cognitively impaired patients. This may be why that although one patient had their best performance on the task, it was by a relatively narrow margin.

Even if a lexical task tends to hold in the presence of cognitive impairment, in order to function as a main predictor of premorbid function, it must also be substantially correlated with some other task in non-clinical participants. This is generally explored with tests of intelligence. In Study 2, we explore this second requirement of premorbid estimation, determining how well the WAT, the SpanLex, and the SCIRT associate with intelligence test scores in a non-clinical sample. For this study the WAT-Sentences test was not included. We also used simulated cognitively impaired performance in this control sample to explore how different combinations of reading tasks might predict IQ scores.

Study 2: Correlations Between Lexical Tasks and IQ

Method

Design and Participants

The aim of this study was to explore the correlations between lexical ability and intelligence test scores. However, following the observations in Study 1 that there may be inter-individual variation in which lexical tasks are most affected, we explored this further. This was achieved by simulating focal impairments by selectively reducing scores and examining the effects on the ability of lexical tasks to predict IQ scores. To this end, we included the 12 elderly controls from Study 1 and recruited an additional 94 control participants (aged 18-65). All were screened for developmental and acquired neurological illness. The whole sample of 106 participants had a mean age in years of 37.73 (SD = 19.50, range 18 - 93), and 58 (55%) were male. The mean years of formal education was 13.92 (SD = 3.91).

The first analyses used the full sample, and for later analyses they were randomly split into two groups. One was the Development Sample (*n* = 53), which was used to develop regression formulas to predict IQ scores from lexical task scores (mean age = 37.47, SD = 20.71; 55% male). The other was the Test Sample (*n* = 53) which was used to explore how well the regression formulas function in different contexts (mean age = 37.98, SD = 18.41; 55% male).

Materials

Lexical Reading Tests. As in study 1 we used the stem-completion task (SCIRT; Pluck, 2018) and the single word pronunciation task (WAT; Del Ser et al., 1997). We also used the lexical-decision task (SpanLex; Pluck, 2020), however, in this study we used the full 36-item version.

Test of Intelligence. We used the Spanish-language Wechsler Adult Intelligence Test IV (Wechsler, 2012) in a validated seven-subtest version (Meyers et al., 2013), including Similarities, Block Design, Digit Span, Arithmetic, Information, Coding, and Picture Completion subtests. This was validated and normed in Spain. However, the between-individual variance in performance is likely to be different in Ecuador. To produce a typical intelligence test score distribution, age-adjusted standard scores for each subtest were calculated from the WAIS-IV manual. The standard scores were then converted to *z* scores, summed, and the product transformed to have a mean of 100 and a standard deviation of 15.

Procedure

This data was collected at four different cities in Ecuador (Quito, Manta, Guayaquil and Riobamba). A wide range of SES backgrounds were sampled. All participants provided written informed consent in accordance with the protocol which had been approved by an appropriate research ethics committee. All interviews were conducted one-to-one in private rooms. Upon completion of the tests, each participant was debriefed and thanked for their participation. Each was given US$20 as compensation for their time. Data loss was minimal; however, the WAT was erroneously not performed with one participant, and the SCIRT with one participant.

Results and Discussion of Study 2

The data supporting this study is publicly available at http://dx.doi.org/10.23668/psycharchives.2897. Scores on the three lexical tasks and the IQ test had normal distributions based on analysis of skew and kurtosis. In the full sample, age was significantly and positively correlated with all of the lexical reading tasks (WAT r = .28, p = .004, SpanLex r = .32, p = .001, SCIRT r = .25, p = .010). However, this is likely caused by the combination of the high functioning, and high socioeconomic status, elderly control group from Study 1, with the new sample of participants from mainly lower and middle socioeconomic backgrounds. If those two groups are analyzed separately, in neither group are there significant associations between age and task performance. We have therefore ignored age as factor in later analyses.

Formula development

In the next step we performed three separate simple linear regression analyses, in each, IQ score was the dependent variable, and one of the three lexical tasks was the independent variable. This was limited to the 53 participants in the Development Sample. For the WAT this produced a model predicting IQ scores, with an R of .78 (R^2 = .61). This was significant, $F(1,50) = 78.77$, $p < .001$. The linear equation predicting IQ scores from WAT scores is IQ = 62.81 + 2.083 (WAT). For the SpanLex the model predicting IQ scores had an R of .63 (R^2 = .40), which was significant, $F(1,51) = 33.94$, $p < .001$. The linear equation predicting IQ scores from SpanLex scores is IQ = 54.61 + 1.821 (SpanLex). Finally, for the model predicting IQ scores from SCIRT scores, there was an R of .79 (R^2 = .62) and the model was also significant, $F(1,51) = 82.91$, $p < .001$. The linear equation to predict IQ scores from SCIRT scores is IQ = 60.81 + 3.292 (SCIRT).

Assessment of Prediction Accuracy and Precision

In the next stage we used those three regression equations to estimate IQ scores in the Test Sample, based on performance of the three lexical tasks. These were then compared to the actual observed IQ scores in the Test sample. We also calculated the median of the three different IQ estimates, on the assumption that it may be less affected by outlying scores and may give a better estimate of observed IQ scores than single tests. Scattergrams are provided in Figure 2, with regression lines, showing the how well observed IQ scores were predicted from the four different methods. We were particularly interested in assessing the accuracy of the different methods, that is, how close the predictions are to actual observed performance. Accuracy is summarized as the effect size (Cohen's *d*) of the difference between predicted and actual observed IQ scores. We were also interested in the precision of the errors, in this context, meaning the variability of the deviations from the observed mean. This is summarized by the range of estimated IQ – observed IQ differences.

[Figure 2 near here]

We also experimented with simulated cognitive impairments. Clinical neuropsychologists rarely have access to the actual observed premorbid scores of patients. However, as we measured IQ in a non-clinical sample, we can simulate the effects of cognitive impairments on the lexical tasks, and examine how well they continue to estimate the observed IQ scores that we have. This can be achieved by systematically reducing the scores. We did this in two ways.

Firstly, to simulate focal cognitive impairments we randomly selected 18 participants and reduced their WAT scores by 1.5 standard deviations (from the sample mean). We did the same to the SpanLex scores of a different 18 participants. The remaining 17 had their SCIRT score reduced by the same amount. Thus, all of the participants had some simulated focal cognitive impairment. We then re-estimated the actual observed IQ scores based on the three different lexical reading tasks using the regression formulas developed earlier. We again also produced a median estimated IQ score, that is, for each participant the middle score of the IQ estimates provided by the WAT, SpanLex, and SCIRT. The median was chosen, rather than the mean, as to estimate in the presence of impairment, without having scores from single impaired performances unduly affecting the central tendency estimate.

Secondly, we simulated global cognitive impairments, that nevertheless vary in how much different functions are affected. For the WAT scores, for 18 randomly selected participants their raw score was reduced by 0.25 standard deviations, for another 18 participants, WAT scores were reduced by 0.5 standard deviations, and the remaining 17 had their WAT scores reduced by 1 standard deviation. This procedure was repeated for the SpanLex, and then for the SCIRT. We again used the formulas developed earlier with the Development Sample to estimate IQ scores and compared the estimated scores with the actual observed scores. As previously, we also calculated the median estimated IQ score on those three lexical tasks. The difference between predictions without simulated impairment and the two different simulated impairment patterns can be used to estimate how resistant to impairment the different measures are.

We first examined the correlations between the four different estimators of ‘premorbid’ ability and actual observed IQ. As would be expected, the estimated IQ scores in the Test Sample, based on the formulas from the Development Sample, were highly correlated with actual observed IQ scores. Within the Test Sample, the WAT estimated IQ correlated $r = .70$, the SpanLex correlated $r = .74$, and SCIRT correlated $r = .70$ with observed IQ scores (all $p < .001$). In the Test Sample we now have the opportunity to also test the correlation of the median estimator (i.e., the median of the three reading task IQ estimates). This produced the highest correlation with observed IQ scores, $r = .80$, $p < .001$. Table 2 contains more detailed information on how well the different estimators performed in approximating the actual observed IQ scores. In addition, the table includes details of how they performed under simulated focal impairment and simulated global impairment.

[Table 2 near here]

Coming first to the estimations without any simulated impairment, all measures slightly overestimated the observed IQ scores. Examination of the Cohen’s *d* values gives an idea of the level of misestimation, in this case overestimation. The worst performing lexical reading task for accuracy was the WAT which had the largest effect size, and overestimated IQ by more than four points. The most accurate was the SCIRT, with the lowest effect size, and provided estimates of IQ that were on average within 2 points of the observed IQ. However, it also important to consider the spread of misestimated scores. In Table 2 it can be seen that each of the lexical tasks in some cases estimated scores that were more than 20 points higher or lower than the observed scores. Worst in this respect was the SCIRT which had a range of discrepancies of over 50 IQ points (from lowest to highest misestimation of IQ). However, the median estimated IQ score was the most precise as it had the least extreme misestimations, ranging only 36 points from its lowest to highest misestimation.

Considering the effects of IQ estimation under simulated focal cognitive impairments, the median score outperforms all of the individual lexical tasks in accurately estimating observed IQ scores. It produces the closest mean, and lowest effect size, *d*, between the estimated and observed values. It shows the least sensitivity to cognitive impairment, with less drop from impaired to unimpaired performance, and it has the smallest range of misestimated scores, suggesting higher precision.

Different results are found when examining the IQ estimations made under simulated global cognitive impairment. This time the word pronunciation task, the WAT, was most accurate at the group level in estimating the observed IQ scores, as indicated by the small effect size, *d*. However, the lexical decision task, the SpanLex, actually had the lowest drop caused by the simulated global cognitive impairment. The median of the three different estimated IQ scores over the three lexical tasks again had the smallest range of misestimation, suggesting the best precision.

Clearly, which approach to estimation works best at predicting IQ scores depends on the context. However, in general, it could be argued that the word pronunciation task, the

WAT, performed well, as did the method of taking the median IQ score of the three different methods. As a crude summary, we calculated the mean performances of the different methods of IQ estimation over the three investigated scenarios (no impairment, simulated focal impairment, and simulated global impairment). These are shown in Table 3. From this, we can see that, overall, taking the median score of the lexical tasks had the best precision, accuracy, and resistance to cognitive impairment. This is probably because the veracity of any estimator based on correlation between scores is dependent on the strength of the correlation. Taking the median of the three measures had a higher correlation with observed IQ than any of the single tasks alone. In fact, the correlation of r = .80 is remarkably high, indicating 64% shared variance between the two measures (r^2 = .64). Furthermore, regressions based on single tests can only give a limited number of predicted scores, with equal intervals between scores. This is the slope of the equation. For the WAT, for example, each single point increase in performance is equal to 2.08 IQ points. Taking the median necessarily allows a more diverse spread of estimated IQ scores. This likely contributes to increased precision of estimates.

[Table 3 near here]

We suggest that the best predictions may come from using multiple tasks to estimate premorbid function, and taking the median value of those estimates. In our case we used lexical tasks, but other tasks might function in similar ways, providing they tend to hold in the presence of cognitive impairment. It should be noted that such tasks would not need to consistently hold, but holding sometimes, may be enough.

As suggested earlier, although premorbid estimation of IQ is a valuable tool in clinical neurosciences, intelligence is often not the measure that principally interests clinical psychologists, nor cognitive neuroscientists. The ability to estimate premorbid levels of other functions would considerably increase the usefulness of these techniques. This has been previously recognized, and there are already commercial tests that estimate premorbid memory ability, in addition to IQ (e.g., Wechsler, 2001, 2011). However, the ability to estimate premorbid executive functions, and related skills associated with frontal lobe functioning, would likely be very useful. This is because executive functions are very commonly observed in neurological illness, and are associated with most forms of psychiatric illness (Snyder et al., 2015). Furthermore, performance on executive function tests is typically closely related to general intelligence (Royall & Palmer, 2014), as such we hypothesized that executive task performance would be highly correlated with performance on our lexical tasks (which themselves seem to be strong measures of intelligence). We therefore conducted an exploratory study on how well the methods described here that estimate IQ, could potentially predict a range of executive function assessments. This is described in Study 3.

Study 3: Correlations Between Lexical Tasks and Executive Function Tests

Method

Design and Participants

The aim of this study was to examine the correlation *r* values between lexical task performance and measures of executive functions, including measures of theory of mind ability. As we have demonstrated the 'hold' status of lexical tasks, the remaining issue is the extent that they correlate with other cognitive tests, those which would be expected to be sensitive to neurological or psychiatric illness. In this exploratory study we recruited 37 participants who were working in a university in Quito, or friends and family members of the researchers. The aim was to recruit people from a range of backgrounds, this included three postgraduate mature students, but no undergraduates were included. The mean age was 35.70, SD = 10.27, range = 18 - 60) and 16/37 (43%) were male. All were assessed with a range of cognitive assessments. An additional sample of 23 undergraduate students was also assessed, mean age 21.04 (SD = 1.67), 12/23 male (52%). However, data from participants still within their educational trajectory are not appropriate for these purposes as their cognitive ability is still developing, and they are generally too homogeneous to give a true reflection of correlation strengths. The undergraduate sample was collected as part of a student thesis (Ruales Chieruzzi, 2015) and are not reported here. However, their responses on the Cognitive Estimates Test were included in setting criteria for abnormality. This is described further below.

Materials

Lexical Tests. We used the same three lexical tasks as used in Study 2. That is, the WAT, the SpanLex, and the SCIRT. However, the version of the SpanLex used here was the abbreviated 17-item version, the same as that used in Study 1. In addition, we again calculated the median score of the three lexical tasks and used that as a separate variable to correlate with the executive function tasks. To do this, the distribution of scores from each of the three lexical tests was converted to *z* scores. Then the median *z* score across the three tests, for each participant, was taken to include in the new median variable.

Tests of Executive Function. We included eight tests of executive function that could be described as ‘cool’, being those that are primarily cognitive, and ‘hot’ which also include motivational or emotional processing (Zelazo et al., 2005). The hot tests were on two different aspects of Theory of Mind. A previous study reported that word pronunciation, measured with a version of the NART, did not correlate sufficiently with any neuropsychological tests, other than intelligence, with the criterion being based on prediction accuracy (Schretlen et al., 2005). However, we note that their best performing tests were Trail Making Part B and Cognitive Estimates, both verbal tests. In addition, one group reported a correlation between NART scores and phonemic verbal fluency performance, of $r = .67$ (Crawford et al., 1992), which is close to the levels found for associations with IQ tests. Learning from these, we included only language-based executive function tasks. In general, the current practice of estimating premorbid IQ in clinical contexts, is based on the observation that IQ scores typically correlate at $r > .70$ in non-clinical populations (Bright et al., 2018). Correlation *r* values of that magnitude also indicate that greater than 50% of the variance in the predicted variable is shared with the predictor variable. We therefore took $r > .70$ as a threshold indicating that a lexical task could predict executive test scores at equivalent levels to the current methods of premorbid IQ estimation.

We also followed Schretlen et al. (2005) in adjusting all of the executive function test scores for age via regression analyses, as the premorbid predictions of intellectual ability (i.e., the current standard method) use age-adjusted IQ scores.

Cognitive Estimates. First published by Shallice and Evans (1978), this test is sensitive to frontal lobe damage and measures top-down cognitive control. It involves asking questions which people are highly unlikely to know the answer to, such as: 'what is the average length of a human spine'? They have to then make a guess. We used a set of nine questions from a previous Spanish-language study of cognitive estimation (Zamorano et al., 2011). However, we used the scoring system from a validated and standardized English language version (MacPherson et al., 2014). That scoring system uses percentiles to classify abnormal responses based on their deviation from the mean. The point thresholds were defined within this study with the full sample of 60 participants (23 undergraduates and 37 non-undergraduates). Total scores were then calculated based on sum of the nine questions, each scored from 0-3. Higher scores indicate more impaired performance.

Proverb Interpretation. Impaired performance on deriving abstract interpretations from proverbs is associated with damage to the frontal lobes and is said to reflect executive dysfunction (Murphy et al., 2013). The Delis-Kaplan Executive function System (Delis et al., 2001) includes a Proverb Test in which eight different English proverbs are presented, and for each the participant is asked to say what it means. Scoring is based on separate assessment of accuracy and level of abstraction which are combined into a total achievement score (maximum four points per proverb). We used the same grading criteria. However, we substituted the eight English proverbs for eight Spanish proverbs that we have used in similar research previously (Pluck et al., 2020). The potential score range was from zero to 32, higher scores indicating better performance.

Verbal Fluency. The ability to spontaneously produce words is particularly sensitive to damage to the frontal lobes (Tucha et al., 1999), and deficits appear to reflect problems with higher-order executive control (Kave et al., 2011). We used three different verbal fluency tasks. The first was phonemic fluency. Participants were told a letter of the alphabet and were required to say as many words as possible beginning with that letter, ignoring proper nouns. They were given one minute. The experimenter recorded and counted all unique words. This was done first for the letter ‘o’, and then for the letter ‘s’. The data of interest is the sum of words produced for both letters. Next, we used a semantic fluency task, with words to be produced from the category of ‘land animals’, and then the category of ‘forms of transport’. The phonemic and semantic fluency stimuli have previously been standardized for use in Spanish (Pluck & López-Águila, 2019). The third task used was phonemic-semantic alternating fluency, in which participants were instructed to say three words beginning with the letter ‘p’ and then three musical instruments, and repeat the cycle. This increases the executive demands of the fluency task by requiring extradimensional switching, in addition to verbal initiation (Iudicello et al., 2008). The number of unique words was counted; however, points were deducted for perseverative responses, rule violations and incompletion triads. This scoring system is that of the Frontal/Subcortical Assessment Battery (Rothlind & Brandt, 1993).

Reading Span Test. This is said to measure working memory capacity. Participants are required to read several sentences aloud, remembering the final word of each sentence (Daneman & Carpenter, 1980). This is a complex task involving phonological storage but also tasking several aspects of central executive capability (Whitney et al., 2001). Brain imaging has linked the processes underlying reading span performance to cingulate and lateral prefrontal connectivity (Kondo et al., 2004). We used the sentences in Spanish for the Reading Span Test provided by Elosúa et al. (1996). Partial scoring was used (Conway et al., 2005). The potential score range is from zero to twelve points, higher scores indicating better performance.

Reading the Mind in the Eyes Test (RMET). This is designed as an adult measure of initial stages of theory of mind, in which participants must attribute mental states based on examining only the eye region in photographs (Baron-Cohen et al., 2001). This uses 36 monochrome photographs and for each there are four choices of mental state. These are presented visually and read aloud by the experimenter. It is available in Spanish from the test developer (www.autismresearchcentre.com). One point is awarded for each mental state correctly identified. Higher scores indicate better performance.

Faux Pas Test. This is also a measure of theory of mind, but involves participants understanding when somebody does or says something tactless or that causes embarrassment. It requires more advanced theory of mind ability than the RMET with both cognitive and empathic aspects, and is sensitive to frontal lobe impairment (Stone et al., 1998). A Spanish version is available from the test developer (www.autismresearchcentre.com). Twenty scenarios are presented sequentially, in ten of which somebody makes a faux pas. One point is awarded for each faux pas correctly identified and understood, or rejected. High scores indicate better performance.

Procedure

All participants provided written informed consent in accordance with the ethics committee approved protocol. All assessments were performed in one-to-one interviews in a quiet room. The participants performed all of the lexical and executive functions tasks. Upon completion, they were debriefed and received compensation of US$5.

Results and Discussion of Study 3

The data supporting this study is publicly available at http://dx.doi.org/10.23668/psycharchives.2897. Zero-order correlations were performed between the lexical measures and the eight age-adjusted executive tasks. These are shown in Table 4. In general, there were large correlations between all lexical measures and the executive function measures. Several executive function tests had coefficients greater than our previously set threshold of .70. These are Proverb Interpretation, Alternating Fluency, and Reading Span. This suggests that lexical tasks could be used to estimate premorbid ability on these tests, and as all of them are sensitive to frontal lobe pathology, to estimate the magnitude of acquired impairments of performance. The test of Proverb Interpretation may be of particular interest, as impaired performance following frontal lobe damage appears to be independent of impairments to general intelligence (Murphy et al., 2013; Roca et al., 2010). Thus, a premorbid estimation of proverb interpretation could be made rather independently of estimates of premorbid intelligence, and may allow clearer identification of executive dysfunction.

[Table 4 near here]

It is notable that the highest correlations were with tests of ‘cool’ executive functions. Despite hypothesizing links between lexical task performance and ‘hot’ functions such as theory of mind, our data suggest that any associations were not of sufficient magnitude to support their use in premorbid estimation, at least at the criterion we set. Furthermore, of the verbal fluency tasks, it was phonemic-semantic alternating fluency that correlated most with lexical reading tasks, more so than phonemic fluency or semantic fluency tested alone. This again is of interest, because the alternating version is the most clearly demanding of executive resources, given that it includes an additional switching component, one of the core executive functions (Miyake et al., 2000). Similarly, performance on the Reading Span Test could be potentially predicted based on lexical reading tasks such as those discussed here. Reading span performance is thought to reflect several aspects of central executive efficacy (Whitney et al., 2001) which may, therefore, allow accurate estimation of the extent of acquired working memory impairments. These comparisons could all be very useful, both clinically and in research.

General Discussion

In three studies we have shown that lexical tasks are good predictors of other cognitive abilities, with potential for use clinically as measures of premorbid ability. Although this has been known for some time, the current results suggest some ways that this approach could be improved.

Firstly, from Study 1, we showed that patients with cognitive impairments due to neurological illness are heterogeneous regarding the extent to which lexical task performance is preserved. This is of course fairly obvious, considering the wide range of neuropsychological dissociations of language function that have been described (e.g., Ardila, 2010). Furthermore, the fact was already suggested by a study of dementia patients that word pronunciation held while lexical decision did not (Beardsall & Huppert, 1997), and with a different patient group, the opposite pattern (McFarlane et al., 2006). The contradictions have only mainly been evident when comparing between studies, because past research has tended to compare dementia patients with control participants at the group level, ignoring heterogeneity of performance within the groups. In our case-series analysis, we found that although single-word pronunciation was the most preserved ability in our sample, in a minority of patients it was the most impaired. This implies that relying on a single ability such as word pronunciation to estimate premorbid intelligence is unreliable. Indeed, we confirmed in Study 2 that the use of three different tasks tapping different lexical skills performs before better than any individual test, if the median value is taken as the premorbid estimate. Although this may seem cumbersome, all of the tests are rapidly administered, and in our study could be applied in about 12-15 minutes. Considering the importance of the premorbid estimation in defining the patient's overall cognitive profile, this seems reasonable. We chose the median value as a measure of central tendency, rather than the mean, as it less affected by individual scores that deviate from the others in the set (such as when a patient has a focal cognitive impairment depressing scores on only one test).

A weakness of the current approach, taking the median of three separate tasks, is that all rely on lexical knowledge. The reason we took this approach is that lexical tasks are the only cognitive tests that are known to frequently ‘hold’ in the presence of cognitive impairments. However, their conceptual similarity as lexical tasks means that they sample a narrow-range of cognitive abilities, making them vulnerable to being simultaneously impaired. Ideally, a wider range of ‘hold’ tests should be identified in future research. Nevertheless, our proposal to use the median of several tests remains valid. We should note however, that others have also suggested that application of multiple tests may be better than relying on single measures. For example, Vanderploeg et al. (1996) have argued that the best performance from a range of cognitive tests may be an appropriate method to increase accuracy of estimates of premorbid cognitive function.

Although in this research we focused on the median score from three different assessments, further studies could explore whether the mean or median are the best estimators of central tendency in clinical practice. Furthermore, machine learning methods, where the best relative weights of multiple tests, or even the best items from within tests, have recently been applied to neuropsychology to improve prediction accuracy (Johnson et al., 2014; van der Linde & Bright, 2018). Such computational methods could potentially improve the accuracy of the methods described here.

In our third study we showed that this our approach of taking the median of three different lexical tasks also applies to neuropsychological tests associated with frontal lobe impairments. This also extends the potential ability of lexical tasks as premorbid estimators. Frontal lobe impairment and associated dysexecutive symptoms are a substantial cause of problems in activities of daily living (Godbout et al., 2005; Putcha & Tremont, 2016). However, intelligence tests have frequently been shown to be insensitive to the cognitive impairments associated with such damage (e.g., Blair & Cipolotti, 2000; Cato et al., 2004). Thus, the ability to make premorbid estimations of executive test scores, rather than intelligence, could be particularly useful in neuropsychological assessment. Our results suggest that working memory, abstract reasoning and generation/switching impairments could be estimated for premorbid levels. It is true that all tests that correlate at least somewhat could potentially be used to predict each other. However, our results suggest that verbal working memory (Reading Span Test), abstract reasoning (Proverb Test) and verbal fluency (Alternating Fluency test) could be predicted at levels of accuracy similar to those achieved by current methods employed to estimate premorbid IQ scores.

An important caveat is that, as the tasks used were all of word knowledge, the methods will not be suitable for all patients. Single word reading and pronunciation tasks, such as the WAT, cannot be applied to people with dysarthria, and none of the tasks would be appropriate for patients with dyslexia. The SCIRT would not be appropriate for patients with semantic dementia. In general, as with the similar tests currently in common clinical use such as the NART (Nelson & Willison, 1991), WTAR (Wechsler, 2001), and TOPF (Wechsler, 2011), they could only be applied clinically when there is no suspicion of visual or language disorder.

Our research is merely suggestive, and does not supply any tools that can be immediately applied anyway. However, the research provides some new experimental approaches: use of case-series instead of group comparisons, stem-completion naming, simulated cognitive impairments, and the use of median scores from multiple tests represent new ways to explore premorbid estimation. These could all be further investigated, potentially leading to the development of novel assessment protocols for quantifying the magnitude of acquired cognitive impairments. Even small improvements may be important, as premorbid estimation provides a cornerstone of the overall assessment in clinical neuropsychology and behavioral neurology.

Some limitations of the current research should be acknowledged. Our sample of neurological patients in Study 1 is relatively small, and thus results should be considered preliminary. Replication with a larger sample is warranted, as the patterns reported here may be not be found in a larger data set. Also, the sample was quite heterogeneous with regard to etiology of the cognitive impairments, which could be considered a further limitation. Many previous studies of this type have included patients with only a single, well-defined neurological illness (though see Beardsall & Huppert, 1997 and Vanderploeg et al., 1996 for exceptions). However, in actual practice, assessments for premorbid function are applied across a range of disorders involving cognitive impairment. From this perspective, inclusion of a heterogeneous group could be seen as making the research more relevant to clinical practice. A further limitation is that we did not assess the healthy participants for neurological disorders or mild cognitive impairment, other than screening based on a brief neurological history. We can not therefore be sure that all of our non-clinical participants were neurological healthy. Our method of simulating cognitive impairment in the assumed healthy participants in Study 2, by reducing observed test scores by set amounts, is an imperfect method of studying the effects of select cognitive impairments on other test scores. We suggest that results from that method be interpreted cautiously.

It should be further acknowledged that the current research was conducted in Spanish, and in South America (Ecuador), and some of the lexical tests were experimental, although we have previously demonstrated their psychometric properties. There is high variability in cognitive skills in Ecuador, linked to extremes in socioeconomic privilege (Pluck et al., 2021). This may have enhanced correlations, which may be found to be not as high in more developed countries. However, Ecuador is not unusually in its socioeconomic diversity, from a global perspective.

In conclusion, lexical tests are good predictors of premorbid cognitive functions. The use of multiple techniques to estimate premorbid function could be explored further, as could the prediction of test scores other than intelligence. This could enhance the fidelity of cognitive assessments, with benefits to practice and investigation in clinical neurosciences.

Geolocation information: Data collection was at four different sites in Ecuador.

Quito: 0.197191°S, 78.436196°W

Manta: 0.942285°S, 80.733242°W

Guayaquil: 2.182906°S, 79.897707°W

Riobamba: 1.680368°S, 78.641198°W

References

Apolinario, D., Brucki, S. M., Ferretti, R. E., Farfel, J. M., Magaldi, R. M., Busse, A. L., & Jacob-Filho, W. (2013). Estimating premorbid cognitive abilities in low-educated populations. *PLoS One, 8*(3), e60084. https://doi.org/10.1371/journal.pone.0060084

Ardila, A. (2010). A proposed reinterpretation and reclassification of aphasic syndromes. *Aphasiology, 2*(3), 363-394. https://doi.org/10.1080/02687030802553704

Arroyo-Anlló, 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. https://doi.org/10.1159/000347223

Baddeley, A., Emslie, H., & Nimmo-Smith, I. (1992). *The Speed and Capacity of Language-Processing Test*. Thames Valley Test Company.

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. https://doi.org/10.1037/a0012919

Baron-Cohen, S., Wheelwright, S., Hill, J., Raste, Y., & Plumb, I. (2001). The "Reading the Mind in the Eyes" Test revised version: a study with normal adults, and adults with Asperger syndrome or high-functioning autism. *The Journal of Child Psychology and Psychiatry, 42*(2), 241-251. https://doi.org/10.1017/S0021963001006643

Beardsall, L., & Huppert, F. (1997). Short NART, CCRT and Spot-the-Word: comparisons in older and demented persons. *British Journal of Clinical Psychology, 36*(4), 619-622. https://doi.org/10.1111/j.2044-8260.1997.tb01266.x

Beardsall, L., & Huppert, F. A. (1994). Improvement in NART word reading in demented and normal older persons using the Cambridge Contextual Reading Test. *Journal of*

Clinical and Experimental Neuropsychology, *16*(2), 232-242. https://doi.org/10.1080/01688639408402634

Blair, R. J., & Cipolotti, L. (2000). Impaired social response reversal. A case of 'acquired sociopathy'. *Brain, 123*(6), 1122-1141. https://doi.org/10.1093/brain/123.6.1122

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. https://doi.org/10.1080/01690960701527317

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. https://doi.org/10.1080/09602011.2016.1231121

Cato, M. A., Delis, D. C., Abildskov, T. J., & Bigler, E. (2004). Assessing the elusive cognitive deficits associated with ventromedial prefrontal damage: a case of a modern-day Phineas Gage. *Journal of the International Neuropsychological Society, 10*(3), 453-465. https://doi.org/10.1017/S1355617704103123

Conway, A. R., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review*, *12*(5), 769-786. https://doi.org/10.3758/BF03196772

Conway, S. C., & O'Carroll, R. E. (1997). An evaluation of the Cambridge Contextual Reading Test (CCRT) in Alzheimer's disease. *British Journal of Clinical Psychology, 36*(4), 623-625. https://doi.org/10.1111/j.2044-8260.1997.tb01267.x

Crawford, J. R., Millar, J., & Milne, A. B. (2001). Estimating premorbid IQ from demographic variables: a comparison of a regression equation vs. clinical judgement. *British Journal of Clinical Psychology, 40*(1), 97-105. https://doi.org/10.1348/014466501163517

Crawford, J. R., Moore, J. W., & Cameron, I. M. (1992). Verbal fluency: a NART-based equation for the estimation of premorbid performance. *British Journal of Clinical Psychology, 31*(3), 327-329. https://doi.org/10.1111/j.2044-8260.1992.tb00999.x

Crawford, J. R., Stewart, L. E., Cochrane, R. H. B., Parker, D. M., & Besson, J. A. O. (1989). Construct validity of the National Adult Reading Test. *Personality and Individual Differences, 10*(5), 585-587. https://doi.org/10.1016/0191-8869(89)90043-3

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior, 19*(4), 450-466. https://doi.org/10.1016/S0022-5371(80)90312-6

de Oliveira, M. O., Nitrini, R., Yassuda, M. S., & Brucki, S. M. (2014). Vocabulary is an appropriate measure of premorbid intelligence in a sample with heterogeneous educational level in Brazil. *Behavioural Neurology, 2014*, 875960. https://doi.org/10.1155/2014/875960

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. https://doi.org/10.1006/brcg.1997.0877

Delis, D. C., Kaplan, E., & Kramer, J. H. (2001). *Delis-Kaplan Executive Function System*. Psychological Corporation.

Elgh, E., Domellof, M., Linder, J., Edstrom, M., Stenlund, H., & Forsgren, L. (2009). Cognitive function in early Parkinson's disease: a population-based study. *European Journal of Neurology, 16*(12), 1278-1284. https://doi.org/10.1111/j.1468-1331.2009.02707.x

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

Gil, G., Magaldi, R. M., Busse, A. L., Ribeiro, E. S., Brucki, S. M. D., Yassuda, M. S., Jacob-Filho, W., & Apolinario, D. (2019). Development of a word accentuation test for predicting cognitive performance in Portuguese-speaking populations. *Arquivos de Neuro-Psiquiatria, 77*(8), 560-567. https://doi.org/10.1590/0004-282x20190089

Godbout, L., Grenier, M. C., Braun, C. M., & Gagnon, S. (2005). Cognitive structure of executive deficits in patients with frontal lesions performing activities of daily living. *Brain Injury, 19*(5), 337-348. https://doi.org/10.1080/02699050400005093

Green, R. E., Melo, B., Christensen, B., Ngo, L. A., Monette, G., & Bradbury, C. (2008). Measuring premorbid IQ in traumatic brain injury: an examination of the validity of the Wechsler Test of Adult Reading (WTAR). *Journal of Clinical and Experimental Neuropsychology, 30*(2), 163-172. https://doi.org/10.1080/13803390701300524

Heirene, R., John, B., & Roderique-Davies, G. (2018). Identification and evaluation of neuropsychological tools used in the assessment of alcohol-related cognitive impairment: a systematic review. *Frontiers in Psychology, 9*, 2618. https://doi.org/10.3389/fpsyg.2018.02618

Iudicello, J. E., Woods, S. P., Weber, E., Dawson, M. S., Scott, J. C., Carey, C. L., Grant, I., & Group, H. I. V. N. R. C. (2008). Cognitive mechanisms of switching in HIV-associated category fluency deficits. *Journal of Clinical and Experimental Neuropsychology, 30*(7), 797-804. https://doi.org/10.1080/13803390701779578

Jensen, A. R. (2001). Vocabulary and general intelligence. *Behavioral and Brain Sciences, 24*(6), 1109-1110. https://doi.org/10.1017/S0140525X01280133

Johnson, P., Vandewater, L., Wilson, W., Maruff, P., Savage, G., Graham, P., Macaulay, L. S., Ellis, K. A., Szoeke, C., Martins, R. N., Rowe, C. C., Masters, C. L., Ames, D., & Zhang, P. (2014). Genetic algorithm with logistic regression for prediction of progression to Alzheimer's disease. *BMC Bioinformatics, 15 Suppl 16*, S11. https://doi.org/10.1186/1471-2105-15-S16-S11

Kave, G., Heled, E., Vakil, E., & Agranov, E. (2011). Which verbal fluency measure is most useful in demonstrating executive deficits after traumatic brain injury? *Journal of Clinical and Experimental Neuropsychology, 33*(3), 358-365. https://doi.org/10.1080/13803395.2010.518703

Kim, H. Y. (2013). Statistical notes for clinical researchers: assessing normal distribution (2) using skewness and kurtosis. *Restorative Dentistry and Endodontics, 38*(1), 52-54. https://doi.org/10.5395/rde.2013.38.1.52

Kondo, H., Morishita, M., Osaka, N., Osaka, M., Fukuyama, H., & Shibasaki, H. (2004). Functional roles of the cingulo-frontal network in performance on working memory. *Neuroimage, 21*(1), 2-14. https://doi.org/10.1016/j.neuroimage.2003.09.046

Lawrence, E. S., Coshall, C., Dundas, R., Stewart, J., Rudd, A. G., Howard, R., & Wolfe, C. D. (2001). Estimates of the prevalence of acute stroke impairments and disability in a multiethnic population. *Stroke, 32*(6), 1279-1284. https://doi.org/10.1161/01.str.32.6.1279

MacPherson, S. E., Wagner, G. P., Murphy, P., Bozzali, M., Cipolotti, L., & Shallice, T. (2014). Bringing the cognitive estimation task into the 21st century: normative data on two new parallel forms. *PLoS One, 9*(3), e92554. https://doi.org/10.1371/journal.pone.0092554

Martínez-Nicolás, I., Carro, J., Llorente, T. E., & Meilán, J. J. G. (2019). The deterioration of semantic networks in Alzheimer’s disease. In T. Wisniewski (Ed.), *Alzheimer’s disease* (pp. 179-191). Condon. https://doi.org/10.15586/alzheimersdisease.2019.ch11

Matsuoka, K., Uno, M., Kasai, K., Koyama, K., & Kim, Y. (2006). Estimation of premorbid IQ in individuals with Alzheimer's disease using Japanese ideographic script (Kanji) compound words: Japanese version of National Adult Reading Test. *Psychiatry and Clinical Neurosciences, 60*(3), 332-339. https://doi.org/10.1111/j.1440-1819.2006.01510.x

McFarlane, J., Welch, J., & Rodgers, J. (2006). Severity of Alzheimer's disease and effect on premorbid measures of intelligence. *British Journal of Clinical Psychology, 45*(4), 453-463. https://doi.org/10.1348/014466505X71245

Meyers, J. E., Zellinger, M. M., Kockler, T., Wagner, M., & Miller, R. M. (2013). A validated seven-subtest short form for the WAIS-IV. *Applied Neuropsychology: Adult, 20*(4), 249-256. https://doi.org/10.1080/09084282.2012.710180

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: a latent variable analysis. *Cognitive Psychology, 41*(1), 49-100. https://doi.org/10.1006/cogp.1999.0734

Morris, P. G., Wilson, J. T., Dunn, L. T., & Teasdale, G. M. (2005). Premorbid intelligence and brain injury. *British Journal of Clinical Psychology, 44*(2), 209-214. https://doi.org/10.1348/014466505X34174

Murphy, P., Shallice, T., Robinson, G., MacPherson, S. E., Turner, M., Woollett, K., Bozzali, M., & Cipolotti, L. (2013). Impairments in proverb interpretation following focal frontal lobe lesions. *Neuropsychologia, 51*(11), 2075-2086. https://doi.org/10.1016/j.neuropsychologia.2013.06.029

Nebes, R. D., Martin, D. C., & Horn, L. C. (1984). Sparing of semantic memory in Alzheimer's disease. *Journal of Abnormal Psychology, 93*(3), 321-330. https://doi.org/10.1037//0021-843x.93.3.321

Nelson, H. E., & Willison, J. (1991). *National Adult Reading Test (NART)*. Nfer-Nelson.

Pluck, G. (2018). Lexical reading ability predicts academic achievement at university level. *Cognition, Brain, Behavior. An Interdisciplinary Journal, 12*(3), 175-196. https://doi.org/10.24193/cbb.2018.22.12

Pluck, G. (2019). Preliminary validation of a free-to-use, brief assessment of adult intelligence for research purposes: the Matrix Matching Test. *Psychological Reports, 122*(2), 709-730. https://doi.org/10.1177/0033294118762589

Pluck, G. (2020). A lexical decision task to measure crystallized-verbal ability in Spanish. *Revista Latinoamericana de Psicología, 52*, 1-10. https://doi.org/10.14349/rlp.2020.v52.1

Pluck, G., Almeida-Meza, P., Gonzalez-Lorza, A., Muñoz-Ycaza, R. A., & Trueba, A. F. (2018). 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 Alcivar, A. M., Ortiz Encalada, P. A., Tello Carrasco, E., Lara, I., & Trueba, A. F. (2019). Modulation of striatum based non-declarative and medial temporal lobe based declarative memory predicts academic achievement at university level. *Trends in Neuroscience and Education, 14*, 1-10. https://doi.org/10.1016/j.tine.2018.11.002

Pluck, G., Córdova, M. A., Bock, C., Chalen, I., & Trueba, A. F. (2021). Socioeconomic status, executive functions, and theory of mind ability in adolescents: Relationships with language ability and cortisol. *British Journal of Developmental Psychology, 39*(1), 19-38. https://doi.org/10.1111/bjdp.12354

Pluck, G., Crespo-Andrade, C., Parreño, P., Haro, K. I., Martínez, M. A., & Pontón, S. C. (2020). Executive functions and intelligent goal-directed behavior: A neuropsychological approach to understanding success using professional sales as a real-life measure. *Psychology & Neuroscience, 13*(2), 158–175. https://doi.org/10.1037/pne0000195

Pluck, G., Lee, K. H., Rele, R., Spence, S. A., Sarkar, S., Lagundoye, O., & Parks, R. W. (2012). Premorbid and current neuropsychological function in opiate abusers receiving treatment. *Drug and Alcohol Dependence, 124*(1-2), 181-184. https://doi.org/10.1016/j.drugalcdep.2012.01.001

Pluck, G., & López-Águila, M. A. (2019). Induction of fear but no effects on cognitive fluency by theta frequency auditory binaural beat stimulation. *Psychology & Neuroscience*, *12*(1), 53-64. https://doi.org/10.1037/pne0000166

Putcha, D., & Tremont, G. (2016). Predictors of independence in instrumental activities of daily living: Amnestic versus nonamnestic MCI. *Journal of Clinical and Experimental Neuropsychology, 38*(9), 991-1004. https://doi.org/10.1080/13803395.2016.1181716

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

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. https://doi.org/10.1111/j.1467-9450.2008.00677.x

Rothlind, J. C., & Brandt, J. (1993). A brief assessment of frontal and subcortical functions in dementia. *The Journal of Neuropsychiatry and Clinical Neurosciences, 5*(1), 73-77. https://doi.org/10.1176/jnp.5.1.73

Royall, D. R., & Palmer, R. F. (2014). "Executive functions" cannot be distinguished from general intelligence: two variations on a single theme within a symphony of latent variance. *Frontiers in Behavioral Neuroscience, 8*, 369. https://doi.org/10.3389/fnbeh.2014.00369

Ruales Chieruzzi, C. B. (2015). *Medición de funciones neuropsicológicas con el uso de tareas de lectura* [Undergraduate Thesis, Universidad San Francisco de Quito]. Quito, Ecuador. https://repositorio.usfq.edu.ec/bitstream/23000/4638/1/114207.pdf

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. https://doi.org/10.1017/S1355617705050939

Shallice, T., & Evans, M. E. (1978). The involvement of the frontal lobes in cognitive estimation. *Cortex, 14*(2), 294-303. https://doi.org/10.1016/s0010-9452(78)80055-0

Sierra Sanjurjo, N., Montanes, P., Sierra Matamoros, F. A., & Burin, D. (2015). Estimating intelligence in Spanish: regression equations with the Word Accentuation Test and demographic variables in Latin America. *Applied Neuropsychology: Adult, 22*(4), 252-261. https://doi.org/10.1080/23279095.2014.918543

Snyder, H. R., Miyake, A., & Hankin, B. L. (2015). Advancing understanding of executive function impairments and psychopathology: bridging the gap between clinical and cognitive approaches. *Frontiers in Psychology, 6*, 328. https://doi.org/10.3389/fpsyg.2015.00328

Sobral, M., Pestana, M. H., & Paúl, C. (2015). The impact of cognitive reserve on neuropsychological and functional abilities in Alzheimer’s disease patients. *Psychology & Neuroscience, 8*(1), 39-55. https://doi.org/10.1037/h0101022

Stone, V. E., Baron-Cohen, S., & Knight, R. T. (1998). Frontal lobe contributions to theory of mind. *Journal of Cognitive Neuroscience, 10*(5), 640-656. https://doi.org/10.1162/089892998562942

Stott, J., Scior, K., Mandy, W., & Charlesworth, G. (2017). Dementia screening accuracy is robust to premorbid IQ variation: evidence from the Addenbrooke's Cognitive Examination-III and the Test of Premorbid Function. *Journal of Alzheimer's Disease, 57*(4), 1293-1302. https://doi.org/10.3233/JAD-161218

Tucha, O. W., Smely, C. W., & Lange, K. W. (1999). Verbal and figural fluency in patients with mass lesions of the left or right frontal lobes. *Journal of Clinical and Experimental Neuropsychology, 21*(2), 229-236. https://doi.org/10.1076/jcen.21.2.229.928

van der Linde, I., & Bright, P. (2018). A genetic algorithm to find optimal reading test word subsets for estimating full-scale IQ. *PLoS One, 13*(10), e0205754. https://doi.org/10.1371/journal.pone.0205754

Vanderploeg, R. D., Schinka, J. A., & Axelrod, B. N. (1996). Estimation of WAIS—R premorbid intelligence: Current ability and demographic data used in a best-performance fashion. *Psychological Assessment, 8*(4), 404-411. https://doi.org/10.1037/1040-3590.8.4.404

Wechsler, D. (1999). *The Wechsler Abbreviated Scale of Intelligence professional manual*. Psychological Corporation.

Wechsler, D. (2001). *Wechsler Test of Adult Reading*. Psychological Corporation.

Wechsler, D. (2011). *Test of Premorbid Functioning (TOPF)*. The Psychological Corporation.

Wechsler, D. (2012). *Escala de Inteligencia de Wechsler Para Adultos-IV*. Pearson.

Whitney, P., Arnett, P. A., Driver, A., & Budd, D. (2001). Measuring central executive functioning: what's in a reading span? *Brain and Cognition*, *45*(1), 1-14. https://doi.org/10.1006/brcg.2000.1243

Zamorano, E. R., Silva, A. P., Sandoval, A. C. M., Vélez, T. G., & Harfuch, L. A. (2011). Estimación cognoscitiva en adultos con trastorno por déficit de atención con hiperactividad. *Psiquiatría*, *27*(3), 10-27.

Zelazo, P. D., Qu, L., & Müller, U. (2005). Hot and cool aspects of executive function: Relations in early development. In W. Schneider, R. Schumann-Hengsteler, & B. Sodian (Eds.), *Young children's cognitive development: interrelationships among executive functioning, working memory, verbal ability, and theory of mind* (pp. 71-93). Psychology Press.

Table 1

Relative performance, compared to the control group, by individual cases in the cognitive impairment group on the four different lexical tasks used in Study 1

Patient	Diagnosis	WAT	WAT-Sentences	SpanLex	SCIRT
1	VaD	**-1.127**	-5.965	-3.269	-3.484
2	AD	**.408**	.261	-2.605	-3.920
3	AD	-.513	**.261**	.053	-.436
4	AD	-.206	.261	**.718**	-1.742
5	AD	**.715**	-.518	.053	-.871
6	AD	**-.820**	-3.630	-3.934	-4.355
7	FTD	.715	1.039	.718	**2.178**
8		**.101**	-2.852	-3.269	-3.920
9	AD	**.408**	-.128	-1.276	-2.613
10	AD	**1.022**	-3.241	-1.940	-1.307
11	AD	-.820	-1.685	-1.940	**-.436**
12		**.408**	-.128	.053	-2.613
13	MCI	**1.636**	.650	1.382	-.436
14	AD	-3.276	-3.241	**-1.276**	-1.742
15	PD	1.022	1.039	.718	**1.742**

Note. Best performance for each patient, defined as the highest *z* score based on the distribution of scores in the control group, is indicated in bold. For two cases no diagnosis other than ‘senile dementia’ was available.

Table 2

Performance of the different lexical measures in the prediction of observed IQ scores under normal conditions and simulated cognitive impairments in Study 2

Lexical measure	n	Observed mean IQ (SD)	Estimated mean IQ (SD)	Mean difference (SD)	Lowest – highest discrepancy	Range of discrepancies	*p*	*d* [a]	Effect of impairment on estimated scores [b]
Unimpaired									
WAT	53	98.69 (14.43)	102.86 (10.91)	-4.17 (10.30)	-20.32 – 25.37	45.69	.005	0.33	
SpanLex	53	98.69 (14.43)	101.20 (10.57)	-2.51 (9.74)	-24.21 – 20.88	45.09	.067	2.00	
SCIRT	52	98.85 (14.52)	100.634 (12.66)	-1.79 (10.67)	-27.96 – 22.92	50.88	.233	**0.13**	
Median	53	98.69 (14.43)	102.23 (10.00)	-3.54 (8.83)	-18.54 – 17.49	**36.03**	.005	0.29	
Simulated focal cognitive impairment									
WAT	53	98.69 (14.43)	97.61 (13.60)	1.08 (14.03)	-20.32 – 33.48	53.80	.578	0.07	-5.25
SpanLex	53	98.69 (14.43)	95.83 (13.58)	2.86 (11.39)	-22.45 – 33.81	56.26	.073	0.21	-5.36
SCIRT	52	98.85 (14.521)	94.60 (14.30)	4.25 (14.18)	-27.96 – 33.62	61.58	.036	0.30	-6.03
Median	53	98.69 (14.43)	98.13 (10.88)	**0.56** (9.72)	-16.68 – 25.37	**42.06**	.675	**0.04**	**-4.10**

Lexical measure	n	Observed mean IQ (SD)	Estimated mean IQ (SD)	Mean difference (SD)	Lowest – highest discrepancy	Range of discrepancies	p	d [a]	Effect of impairment on estimated scores [b]
Simulated global cognitive impairment									
WAT	53	98.69 (14.43)	96.71 (11.16)	**1.98** (10.54)	-17.59 – 30.83	48.42	.177	**0.16**	-6.15
SpanLex	53	98.69 (14.43)	95.11 (10.76)	3.58 (10.06)	-21.57 - 28.06	49.63	.012	0.28	**-6.08**
SCIRT	52	98.85 (14.52)	93.27 (13.32)	5.58 (11.28)	-24.80 – 35.58	60.38	.001	0.40	-7.37
Median	53	98.69 (14.43)	95.82 (10.89)	2.87 (8.74)	-15.15 – 28.06	**43.21**	.020	0.23	-6.41

[a] mean difference / pooled standard deviation, [b] the difference between lexical task predicated IQ and lexical task predicted IQ with simulated cognitive impairment. The best performing measures in each of the 3 different simulations in terms of accuracy, precision and resistance to impairment are indicated in bold.

Table 3

Summary mean scores of the statistics relating to precision, accuracy, and resistance to cognitive impairment, of the different IQ estimation over the three different scenarios in Study 2

Lexical measure	Range of deviations	Cohen's *d*	Effect of impairment on estimated IQ scores
WAT	49.30	0.03	5.70
SpanLex	50.33	0.10	5.36
SCIRT	57.62	0.19	6.03
Median	40.43	0.01	4.10

Table 4

Correlation r values between lexical task scores and executive function task scores from Study 3

Executive test	WAT	SpanLex	SCIRT [a]	Median
Cognitive Estimates [a]	-.53***	-.61***	-.61***	-.62***
Proverb Interpretation [a]	.68***	**.70*****	**.71*****	**.72*****
Phonemic Fluency	.49**	.59***	.63***	.56***
Semantic Fluency	.46**	.47**	.57***	.49**
Alternating Fluency	.63***	**.79*****	.66***	**.76*****
Reading Span	.68***	.68***	.64***	**.73*****
RMET	.55***	.53**	.60***	.64***
Faux Pas [a]	.49**	.52**	.36*	.45**

Note. All correlation coefficients are Pearson *r* values, except [a] = Spearman's rho due to nonnormal data distributions

**p* < .050, ***p* < .010, ****p* < .001 (two-tailed). *r* values > .70 are indicated in bold.

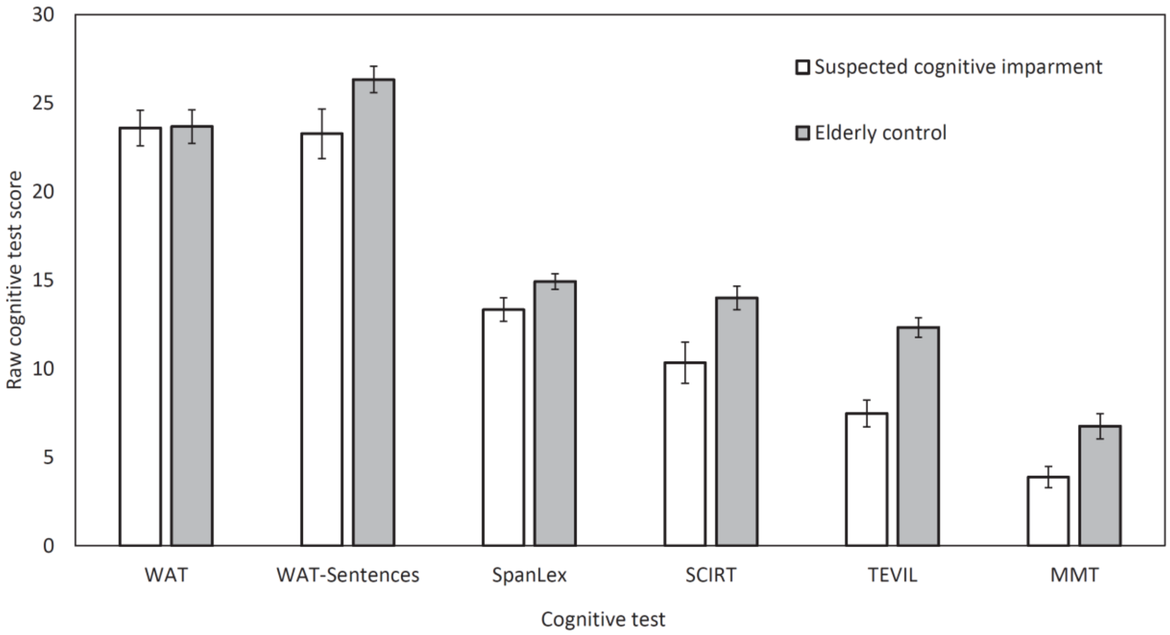

Figure 1

Mean scores (+ SEM) for the cognitively impaired and elderly control group for all cognitive measures in Study 1

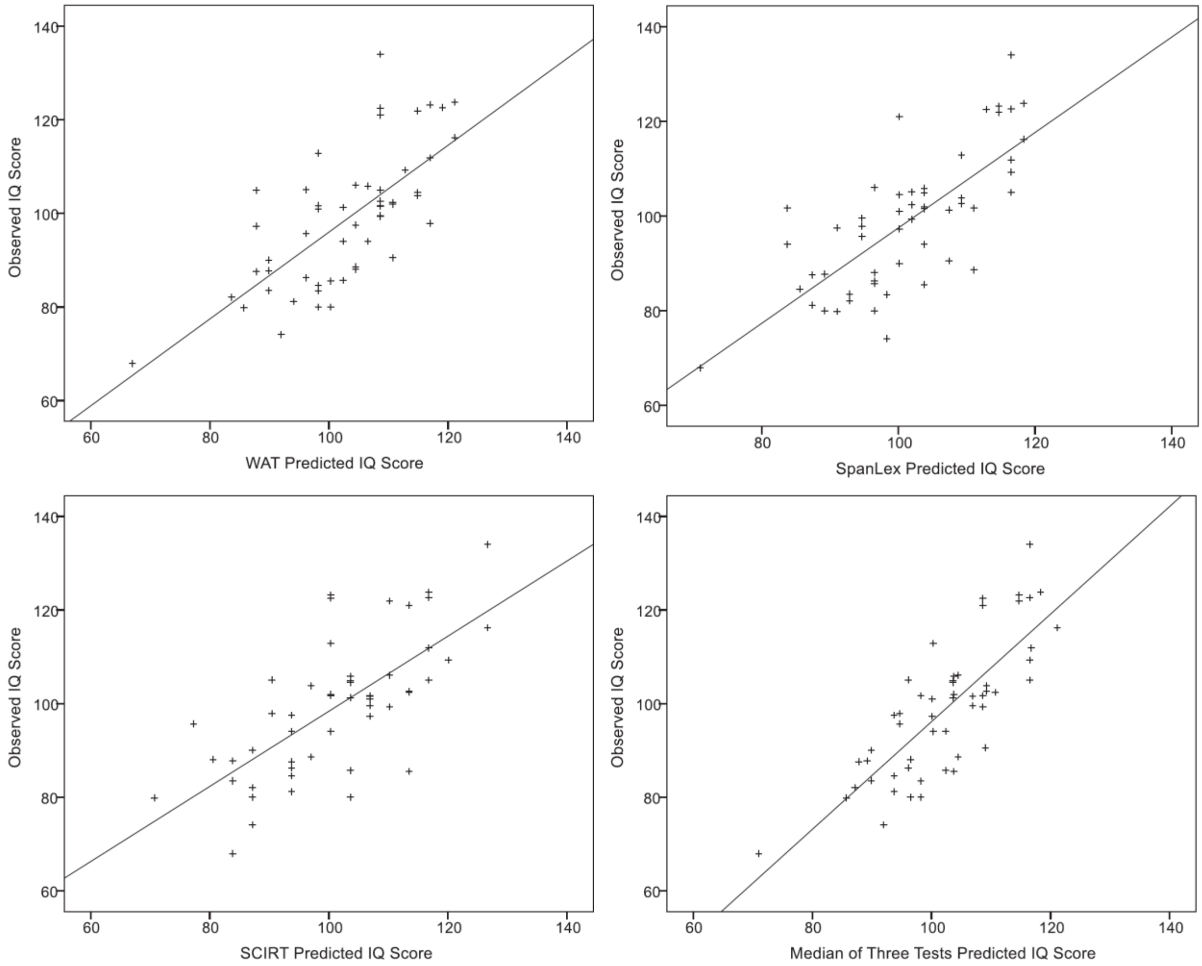

Figure 2

Scattergrams plus regression line for the prediction of IQ scores by the four different methods in Study 2