The Hopkins Verbal Learning Test: An In-depth Analysis of Recall Patterns

Rachel Grenfell-Essam[1], Eef Hogervorst[1, 2, 3], and Tri Budi W Rahardjo[2, 3]

[1]School of Sport, Exercise and Health Sciences, Loughborough University

[2]Centre for Ageing Studies, Universitas Indonesia

[3]Universitas Respati Indonesia

Correspondence to:

Rachel Grenfell-Essam,

School of Sport, Exercise and Health Sciences,

Loughborough University,

Loughborough,

Leicestershire,

LE11 3TU,

UK

Telephone: +44 1509 223497 or E-mail: R.E.Grenfell-Essam@lboro.ac.uk

Abstract

One of the earliest signs of dementia is memory issues and verbal word lists, such as the Hopkins Verbal Learning Test (HVLT), are successfully used for screening. To gain insight in how memory is affected in dementia, and to further improve the efficacy of the HVLT, in-depth analysis of the recall patterns of dementia cases and controls was conducted. Dementia cases and controls were matched for factors that can affect performance, such as age, gender and education level. Word frequency, syllable length, and orthographic neighbourhood size did not differ in the Indonesian version of the HVLT, nor did these characteristics affect recall. However dementia cases showed consistent and poor recall across the three trials; with the worst recall for the 'human shelter' category and best recall for the 'animals' category. Dementia cases also showed impaired accessibility of all categories with reduced subsequent recall from accessed categories and reduced primacy and recency levels. Finally, dementia cases exhibited lower levels of re-remembering and recalling new words, and higher levels of immediate forgetting and never recalling words. It was concluded that utilising the extra information provided by the in-depth analyses of the recall patterns could be beneficial to improve dementia screening.

198 words (max allowed is 200 words)

Keywords: Dementia; Hopkins Verbal Learning Test; HVLT; memory; recall patterns

The majority of dementia cases are expected to reside in developing countries (Hogervorst et al., 2011; Prince et al., 2015). Therefore, it is crucial that effective but easily administered, low technology screening tests for dementia are available which ideally could be administered without the need for specialist training or equipment. The Hopkins Verbal Learning Test (HVLT; Brandt, 1991) is one such screening test that we believe holds promise for use in developing countries (Hogervorst et al., 2011; Xu, Rahardjo, Xiao & Hogervorst, 2014).

The HVLT is a short test of verbal memory taking around ten minutes to administer (Brandt, 1991). The test was revised to include a delayed recall trial, which is copyrighted as HVLT-R (Benedict, Schretlen, Groninger & Brandt, 1998). The HVLT has been favourably compared to other dementia tests, such as: the Mini-Mental Status Examination (MMSE) (e.g., de Jager, Schrijnemaekers, Honey & Budge, 2009; Frank & Byrne, 2000; Kuslansky et al., 2004), the California Verbal Learning Test (CVLT) (Lacritz & Cullum, 1998; Lacritz, Cullum, Weiner & Rosenberg, 2001), CogState (de Jager, et al., 2009), subtests of the Wechsler Memory Scale-Revised (Shapiro, Benedict, Schretlen & Brandt, 1999), and the Brief Visuospatial Memory Test–Revised (Shapiro et al., 1999). The HVLT has also been compared to less commonly used tests and has been shown to be more effective in screening for dementia (see Benedict, Schretlen, Groninger, Dobraski & Shpritz, 1996; Shapiro et al., 1999).

The HVLT has been adapted and/or translated for use in multiple countries (e.g., French: Rieu, Bachoud-Lévi, Laurent, Jurion, & Dallabarba, 2006; Spanish: Cherner et al., 2007; Chinese: Shi, Tian, Wei, Miao & Wang, 2012; Indonesian: Hogervorst et al., 2011). It is well tolerated by participants and is suitable for repeated testing due to its six different forms with minimal learning effects and no ceiling effects (Benedict et al., 1998; Krebs, 1994; Rasmusson, Bylsma & Brandt, 1995; Woods et al., 2005). An optimal cut-off score for

3

the HVLT for dementia cases is 16 to 17 words (total of 3 trials) immediately recalled, but this may vary depending on the country (and translation use) where it is used and/or the age of participants (see Xu, Rahardjo, Xiao & Hogervorst, 2014 for a review). The HVLT displays good sensitivity and specificity for dementia in both Western (Friedman, Schinka, Mortimer & Graves, 2002; Hester, Kinsella, Ong & Turner, 2004; Hogervorst et al., 2002) and non-Western countries (Hogervorst et al., 2011; Shi et al., 2012; Xu et al., 2014).

Given the extent of previous research studying the HVLT it is perhaps surprising that research exploring in-depth performance of dementia patients is very limited. A highly similar test, the CVLT (Delis, Kramer, Kaplan, & Ober, 1987), had explored performance over a wide range of in-depth analyses, such as semantic clustering (the degree to which items from the same category are recalled together), primacy and recency levels, learning rates, intrusion errors, etc. However, within the HVLT, the total recall score is the main measure used for assessment. It is useful for quick identification of probable dementia patients (those who fall below the cut-off score), but the total recall score fails to investigate whether there are any subtler differences within the recall of probable dementia patients and controls.

Only a few papers have explored HVLT recall in more depth. The most relevant of these is Schrijnemaekers, de Jager, Hogervorst and Budge (2006) who provided a brief examination of category recall and serial position effects in an Oxfordshire based cohort of cases and controls. Serial position effects refer to the finding that the location of an item within the list affects its likelihood of being remembered. Primacy effects refer to the recall advantage for words at the start of the list, and recency effects refers to the recall advantage for words at the end of the list (e.g., Jahnke, 1965; Murdock, 1962; Roberts, 1972). Schrijnemaekers et al. (2006) explored whether there were differences in the recall of the three HVLT categories for trials 1 and 3 only. They did not find a significant difference for

dementia cases, however, the trend in both trials 1 and 3 was for better recall of the category 'animals' compared to the categories 'semi-precious stones' and 'human shelter'. Dementia cases recall of all three categories was significantly poorer than controls. They also assessed primacy and recency effects via a grouped serial position analysis for trial 3 only. Recall was significantly lower for dementia cases. Controls showed no primacy and recency effects, but dementia cases showed both. This paper provides an interesting, albeit limited, assessment of recall patterns within the HVLT.

Primacy and recency effects on other word lists have been investigated within dementia research. Typically dementia cases show an impaired primacy effect in word list recall, but a relatively preserved recency effect and so this may be diagnostically useful (e.g., Bayley et al., 2000; Carlesimo, Fadda, Sabbadini, & Caltagirone, 1996; Foldi, Brickman, Schaefer, & Knutelska, 2003; Gainotti & Marra, 1994; Howieson et al., 2011; Massman, Delis, & Butters, 1993; Tierney et al., 1994).

Other papers have focused on semantic clustering within the HVLT. This can be assessed if the order in which the words are recalled is recorded (output order). Gaines, Shapiro, Alt and Benedict (2006) compared semantic clustering indexes in healthy US based adults, patients with probable dementia of Alzheimer's type and patients with vascular dementia. Both types of dementia patients showed lower levels of semantic clustering than controls. Lower levels of semantic clustering within the HVLT have also been found with amnestic mild cognitive impairment (aMCI) patients compared to controls (Malek-Ahmadi, Raj & Small, 2011). This tendency for impaired semantic clustering is well-known within dementia patients (e.g., Carlesimo et al., 1998; Delis et al., 1991; Glosser, Gallo, Clark & Grossman, 2002; Hodges et al., 1999; Perri, Carlesimo, Serra, & Caltagirone, 2005; Weingartner et al., 1981).

As well as the aspects discussed above, there are several general characteristics that can affect the ability to remember a word, such as word frequency, word length (number of syllables), age-of-acquisition, and orthographic neighbourhood size. Surprisingly, although the HVLT allows for analysis of each word, no previous papers have investigated the effect of these characteristics on recall of the individual words. Therefore it is currently unknown how these characteristics vary within the HVLT and what effect dementia might have on recall. There is also a lack of research investigating these characteristics within dementia patients' performance in *free* recall tasks.

Word frequency is perhaps the most important characteristic. However, most research conducted on dementia patients has focused on recognition tasks where the low frequency advantage observed in recall tasks dramatically decreases or disappears (Balota, Burgess, Cortese & Adams, 2002; Wilson, Bacon, Fox, Kramer & Kaszniak, 1983). Preserved word (syllable) length effects in memory span (serial recall) have been shown within dementia patients (Belleville, Peretz & Malenfant, 1996; Morris, 1984).

The expected effect of age-of-acquisition is more complex. Within healthy individuals some papers have found better recall for late acquired words (Dewhurst, Hitch & Barry, 1998; Morris, 1981) whilst others have found no effect (Christian, Bickley, Tarka, & Clayton, 1978; Coltheart & Winograd, 1986; Gilhooly & Gilhooly, 1979; Rubin, 1980). Studies investigating age-of-acquisition in dementia patients have focused on word naming, picture naming and lexical decision tasks. They consistently find an advantage for early acquired words on these tasks (e.g., Cuetos, Herrera & Ellis, 2010; Forbes-McKay, Ellie, Shanks & Vennen, 2005; Kremin, Hamerel, Dordain, De Wilde & Perrier, 2000; Ralph, Graham, Ellis & Hodges, 1998; Silveri, Cappa, Mariotti & Puopolo, 2002) which is in line with healthy individuals (e.g., Barry, Morrison, & Ellis, 1997; Carroll & White, 1973; Gerhand & Barry, 1998; Morrison & Ellis, 1995).

Dementia patients have shown improved performance for words with larger numbers of orthographic neighbours within a recognition task using Spanish speakers (Dunabeitia, Marín & Carreiras, 2009). This pattern has been mirrored within the normal population in serial recall tasks (memory span and immediate serial recall) (Allen & Hulme, 2006; Glanc & Greene, 2012; Jalbert, Neath, Bireta & Surprenant, 2011; Jalbert, Neath & Surprenant, 2011; Roodenrys, Hulme, Lethbridge, Hinton, & Nimmo, 2002).

In the current paper we sought to investigate whether these four characteristics varied at the category and individual word level in Form 1 of the HVLT in both English and Indonesian languages[1]. We are the first to explore the effects of word frequency, word length, age-of-acquisition and orthographic neighbourhood size on performance of dementia cases within the HVLT in either English or Indonesian and the first to compare the two languages[2]. Whilst finalising the current paper, a highly relevant paper was published that looked at effects of word frequency, word length, and orthographic neighbourhood size within controls and dementia cases utilising the German version of the CVLT (Hessler, Fischer & Jahn, 2016). They found that word length had no effect on recall for either controls or dementia cases, that word frequency showed an overall advantage for high frequency words for both controls and dementia cases, and that controls displayed better recall with high orthographic neighbourhood size, whereas dementia cases showed better recall with low orthographic neighbourhood size.

In-depth analysis of the HVLT may allow for greater understanding of the nature of the memory impairment that causes reduced recall within dementia cases. As discussed earlier, dementia cases consistently show a reduced ability to utilise semantic information indicating an encoding issue. Furthermore, dementia cases show a large range of issues: impaired performance in both recollection (e.g., free recall) and familiarity (e.g., old/new and

---

[1] The data were collected from Indonesian speaking participants
[2] Indonesian uses the Latin alphabet making direct comparison easier

forced-choice recognition) tasks, a lack of learning over repeated exposure, impaired recall consistency, reduced subjective organisation, increased sensitivity to proactive interference, higher rates of forgetting, greater susceptibility to intrusions, higher rates of false alarms, greater number of perseverative errors, and impaired retention over time (e.g., Algarabel et al., 2009, 2012; Ally, Gold & Budson, 2009; Butters et al., 1983; Butters, Granholm, Salmon, Grant, & Wolfe, 1987; Dallabarba & Wong, 1995; Dannenbaum, Parkinson & Inman, 1988; Delis et al., 1991; Granholm & Butters, 1988; Grober & Buschke, 1987; Koen & Yonelinas, 2014; Kopelman, 1991; Martin, Brouwers, Cox, & Fedio, 1985; Ober, Koss, Friedland, & Delis, 1985; Perri et al. 2005; Weingartner et al., 1981; Weingartner, Grafman, Boutelle, Kaye & Martin, 1983; Wolk, Dunfee, Dickerson, Aizenstein & DeKosky, 2011). These findings indicate widespread issues with encoding, storage, and retrieval.

However, in most cases the issues researched above do not examine performance from trial to trial (except for recall consistency and subjective organisation). Repetition of the same words across three separate trials within the HVLT allows examination of retrieval dynamics is possible via the tracking of each individual word from trial to trial. One can investigate which words are learnt, retained, and forgotten from trial to trial. Typically the *learning* research above uses a calculation of the difference between the number of words recalled in the last immediate trial compared to the first immediate trial and the *forgetting* rates are calculated as the difference between the number of words recalled in a delayed recall trial compared to the final immediate recall trial. Immediate forgetting from trial to trial is synonymous with failure to benefit from repeated exposure. However, these calculations are not sensitive as to specifically which words are being learnt or forgotten from trial to trial within immediate recall. The words recalled in the final immediate recall trial may be completely different words to the ones recalled in the first immediate recall trial, but if the number of words is the same, then no learning will be the reported outcome.

8

A few studies, however, have looked at intertrial performance to assess learning and forgetting with dementia cases. Woodard, Dunlosky and Salthouse (1999) examined intertrial acquisition and consolidation levels within six dementia cases. Utilising the Rey Auditory Verbal Learning test (Rey, 1964) they determined the levels of gained and lost items across adjacent trials. Gained items reflect acquisition (learning) and lost items reflect impaired consolidation (forgetting). Dementia cases had lower levels of learning and higher levels of forgetting suggesting impaired encoding and storage. These measures were also not correlated with each other, indicating that these assess independent processes.

Moulin, James, Freeman and Jones (2004) sought to replicate and extend Woodard et al.'s findings by utilising a larger sample, comparing controls with a MCI group and using a different test (CERAD word learning list; Welsh, Butters, Hughes, Mohs & Heyman, 1991). Like Woodard et al. they found significantly lower levels of gained items and higher levels of lost (forgotten) items for dementia cases, and no correlation between the two measures. In addition they also found that dementia cases were more impaired on both acquisition and consolidation measures compared to MCI cases, indicating that these both decline as dementia severity progresses. Additionally, Genon et al. (2013) utilised this gained and lost access criteria within another test (CVLT), and also confirmed deficient acquisition and consolidation levels with dementia cases whilst also providing the neural correlates.

We are the first to examine the levels of *re-remembered* (recalled in this trial and previously), *new* (recalled in this trial but not previously, akin to *gained access*), *forgotten* (not recalled in this trial but recalled previously, akin to *lost access*) and *never recalled* (not recalled in this trial or previously) words within the HVLT. We also operationalise these calculations in a different way. Both Woodard et al. (1999) and Moulin et al. (2004) based their measures on consecutive trials. This means that a word recalled in trial 1 and 3 but not 2 would be seen as a gained item (*new*) in trial 3 rather than a *re-remembered* item. Therefore

9

we sought to eliminate this by amending our calculation to take into account recall of specific words from all previous trials.

The first author's previous work, looking at the recall patterns in two other immediate memory tasks (immediate free recall and immediate serial recall), has shown how important in-depth analysis can be to advance a field (Grenfell-Essam & Ward, 2012, 2015; Grenfell-Essam, Ward & Tan, 2013, 2017; Ward, Tan & Grenfell-Essam, 2010). We hope that a deeper exploration of the recall patterns of controls and dementia cases within the HVLT may help improve its diagnostic efficacy. We are the first to investigate whether there are differences between word frequency, word length, age-of-acquisition and orthographic neighbourhood size within the 12 words of the HVLT for both English and Indonesian languages[3] and the first to compare the two languages. We also investigated whether there were differences between controls and dementia cases on the total HVLT score, the overall proportion of words recalled (allowing assessment of learning), category recall and accessibility, the effects of word frequency, word length, and orthographic neighbourhood size on recall[4], the levels of primacy and recency (via serial position curves), and the retrieval dynamics across trials (via the levels of re-remembered, new, forgotten and never recalled words). The majority of these recall patterns have not been previously investigated within the HVLT. We build upon a small number of studies that have investigated these areas by providing a more systematic exploration of recall within all trials and by utilising a larger sample.

The main aim of this paper is to improve the efficacy of the HVLT by discovering whether, over a range of performance measures, there are differences between control and dementia cases. These more subtle differences between control and dementia cases, over and

---

[3] Age-of-acquisition values were not available for the Indonesian version
[4] Using the Indonesian values

above the current overall recall measure currently used, may help screen for earlier signs of dementia.

Method

Sample. Data were taken from Hogervorst et al. (2011), which was interested in investigating the validation of the HVLT and MMSE dementia screening tests in Indonesia. The data was from a cross-sectional study involving 719 elderly Indonesian participants from West Java, Central Java and Jakarta (see Hogervorst et al., 2011 and Yesufu, Bandelow, Rahardjo & Hogervorst, 2009 for specific details of the study design). Informed consent had been obtained before study onset and ethical approval had been obtained from a local Indonesian ethical committee, as well as from Loughborough University in the United Kingdom.

Cognitive measures. Three cognitive measures taken from Hogervorst et al. (2011) were used in this study: HVLT (Brandt 1991), MMSE (Folstein, Folstein & McHugh, 1975), and Instrumental Activities of Daily Living (IADL, Lawton & Brody, 1969). The latter two measures were used to classify control and dementia cases.

HVLT: This test consists of three trials; each trial contains the same 12 words taken from three semantic categories (four words from each category). There are six forms to the test; form 1 was used containing four-legged animals (referred hereafter as *animals*), precious stones (referred hereafter as *stones*), and human dwellings (referred hereafter as *shelter*). The words are read out loud by the experimenter at a rate of around one word every two seconds. At the end of each trial participants' immediate free recall of the words is recorded. A total immediate recall measure is obtained by summing the total number of words correctly recalled from each trial. The maximum total immediate recall score possible is 36. An Indonesian version of the HVLT was used (Hogervorst et al., 2011) and as such all analyses

11

unless otherwise stated are based on this Indonesian version. Back-translation and focus groups were used to further validate the test items for local use. Note that Hogervorst et al. (2011) modified some of the words from the *'precious stones'* category to suit local knowledge following a pilot study[5].

MMSE: This test consists of 11 questions that can be used to assess mental status across a range of five aspects of cognitive ability (orientation, registration, attention and calculation, recall, and language). Note that Hogervorst et al. (2011) adapted the questions for local circumstances similar to Ganguli et al.'s (1995) Hindi version. The maximum score possible is 30 and a cut-off score of 24 or less was adopted for dementia cases.

IADL: This test consists of eight activities, such as ability to use a telephone, housekeeping and responsibility for own medications, and scores their functional ability level for each activity. Note that Hogervorst et al. (2011) used a modified version that can be found within their article where the ability to perform each activity was scored from 0 (unable to do) to 2 (independent ability). The maximum score possible for this modified version is 16 and a cut-off score of 9 or less was adopted for dementia cases.

Statistical Analyses. We classified participants in the data set into control or dementia cases based on their scores on the MMSE and IADL. From the data set of 719 participants, dementia cases were selected if they met all of the following three criteria: they had a MMSE score of 24 or less, an IADL score of 9 or less and had a recorded value for the Indonesian version of the HVLT (i.e. no missing data). Control cases were selected if they met all the following three criteria: they had a MMSE score of 25 or more, an IADL score of 10 or more and had a recorded value for the Indonesian version of the HVLT. Using these criteria 216 participants were excluded because they did not match the cut-off criteria for

---

[5] Please see Table 1 for details

either dementia or control based on their MMSE and/or IADL values and a further 31 were excluded because they had missing values for either HVLT, MMSE or IADL. Differences in demographic characteristics (gender, age and education level) of the control and dementia cases were assessed via Chi-square and Mann-Whitney U tests. Due to significant differences between age and education level in the two groups a matched samples approach was adopted. Participants were matched by their gender, education level (exact match) and age (within 5 years). A Mann-Whitney U test confirmed non-significant differences between the two groups in age.

A set of analyses were performed to examine the word frequency, word length, and orthographic neighbourhood size characteristics mentioned in the introduction and methods section for both English and Indonesian languages (see Table 1)[6]. This included a series of one-way analysis of variances (ANOVAs) with category (3 levels: animals, stones and shelter) as the independent variable and also a series of correlations to see whether any of the word characteristics were associated with each other. An independent sample t-test examined whether the HVLT total recall score differed between dementia cases and controls. The overall proportion of words recalled across the three trials of the HVLT was examined via a two-way mixed ANOVA. A set of analyses were performed to examine category recall and accessibility; this included a series of two-way mixed ANOVAs and also a series of chi-squares and independent samples t-tests. The relationship of word characteristics and recall was then assessed via a series of correlations. Serial positions curves, that allow assessment of primacy and recency within trials, were examined via a series of two-way mixed ANOVAs. Finally, the levels of learning and forgetting were investigated via a series of two-way mixed ANOVAs.

---

[6] This is the only set of analyses which investigate the English version of the HVLT.

For all two-way and three-way ANOVA analyses one of the independent variables was always group (dementia or control) and the results are reported within tables for clarity. Mauchly's Test of Sphericity was checked for all ANOVA analyses containing within-subject independent variables with more than two levels. Where Sphericity was violated the Greenhouse-Geisser correction was applied. Where significant main effects for within-subject variables occurred these were followed up via Bonferroni corrected pairwise comparisons. Where significant two-way interactions occurred these were followed up via simple effects using a Bonferroni correction. An alpha value of 0.05 was used for all statistical analyses unless otherwise stated and SPSS 23.0 was used.

Results

The data was considered in eight different analyses: participant demographics, word characteristics, HVLT total recall score, overall proportion of words recalled, category recall and accessibility, relationship of word characteristics and recall, serial position curves, and retrieval dynamics across trials.

Participant demographics. The first analysis investigated the demographic characteristics of the dementia and control groups. The criteria outlined in the method resulted in 76 dementia cases and 396 control cases. A Chi-square indicated that there was a non-significant difference for gender between the dementia group (Males = 32, Females = 44) and control group (Males = 140, Females = 256), χ^2 (1) = 1.26, p = 0.263. However, Mann-Whitney U tests indicated that the dementia group (mean = 76.5, SD = 8.11) was significantly older than the control group (mean = 67.2, SD = 6.34), Z = -8.56, $p < 0.001$, and that the dementia group had a significantly lower education level than the control group, Z = -7.82, $p < 0.001$.

14

Due to the significant differences in age and education level between the control and dementia cases, a matched samples approach was adopted. Control and dementia participants were matched on gender (exact matches only), education level (exact matches only) and age (matches exact where possible but no further than 5 years apart). This resulted in 62 matched participants in each group (Males = 23, Females = 39). A Mann-Whitney U test indicated that there was no significant difference in age between the dementia group (mean = 74.6, SD = 7.38) and the control group (mean = 73.5, SD = 6.47), Z = -.702, p = 0.483. Full details of the demographic information of the participants can be found in Appendix 1.

Word characteristics. This analysis explored whether the characteristics discussed in the introduction (word frequency, word length, age-of-acquisition, and orthographic neighbourhood size) vary within Form 1 of the HVLT in both English and Indonesian. Table 1 shows the word frequency, word length (number of syllables), age-of-acquisition[7], and orthographic neighbourhood size (based on Coltheart's N; Coltheart, Davelaar, Jonasson, & Besner, 1977) values for the words and categories in Form 1. For the English version: word frequency values represent number of occurrences per million words taken from Brysbaert and New (2009) which are based on the SUBTL database; age-of-acquisition values were taken from Kuperman, Stadthagen-Gonzalez and Brysbaert (2012); and orthographic neighbourhood size were based on Coltheart's N values from the CLEARPOND database (Marian, Bartolotti, Chabal & Shook, 2012). For the Indonesian version: word frequency values were taken from Hermit Dave (2016) and are based on the OpenSubtitles2016 data for Indonesian subtitles[8]. Orthographic neighbourhood size values were calculated by the first author by systematically substituting each letter individually with all possible letters of the

---

[7] No age-of-acquisition values within the Indonesian language were available
[8] These provide raw scores rather than values per million. This word frequency source was the only data set which provided values for all words within the Indonesian version of the HVLT

alphabet to make all possible combinations of orthographic neighbours. These words were then verified as real Indonesian words via at least three separate sources (a translation tool, an Indonesian dictionary and an Indonesian thesaurus).

--Table 1 about here--

Table 1 shows that the individual words display variation within the characteristics for both the English and Indonesian languages. In particular the word frequency values show a large range in both languages. These values cannot be directly numerically compared due to differences in measurement but they appear to follow similar trends. This was confirmed by a significant correlation, $r(12) = 0.941$, $p < 0.001$. Word length shows much lower variability, but more similarity across the two languages (English 1-3 syllables; Indonesian 2-4 syllables) with only 1 word in both languages having the highest number of syllables. However, there was a non-significant correlation between the two languages for word length, $r(12) < 0.001$, $p = 1.000$. Age-of-acquisition values were only available for English but these displayed a wide range of 5.34 years. Orthographic neighbourhood size shows a wider range for English (English 0-24 neighbouring words; Indonesian 0-11 neighbouring words) but they appear to follow similar trends as confirmed by a significant correlation, $r(12) = 0.796$, $p = 0.002$.

To investigate whether each characteristic varied at the category level, a series of one-way ANOVAs were conducted on each characteristic with category (3 levels: animals, stones and shelter) as the independent variable. For the English language these ANOVAs revealed non-significant differences between: frequency of occurrences per million, $F(2, 9) = 1.25$, $MSE = 1119.0$, $p = 0.331$; word length, $F(2, 9) = 1.93$, $MSE = 0.389$, $p = 0.201$; and orthographic neighbourhood size values, $F(2, 9) = 3.16$, $MSE = 49.4$, $p = 0.091$. However, there was a significant difference between the age-of-acquisition, $F(2, 9) = 14.6$, $MSE =$

1.18, p = 0.002. Bonferroni corrected pairwise comparisons showed that animals were acquired significantly earlier than both stones (p = 0.001) and shelter (p = 0.038), but that stones and shelter were not different (p = 0.147). Additionally only two of the characteristics significantly correlated with each other: word length and orthographic neighbourhood size, r (12) = -0.642, p = 0.024. All other correlations were not significant; word frequency and word length, r (12) = -0.123, p = 0.703, word frequency and age-of-acquisition, r (12) = -0.441, p = 0.151, word frequency and orthographic neighbourhood size, r (12) = -0.020, p = 0.950, word length and age-of-acquisition, r (12) = 0.484, p = 0.111, and age-of-acquisition and orthographic neighbourhood size, r (12) = -0.335, p = 0.287.

For the Indonesian language these ANOVAs exploring category differences also revealed non-significant differences between: word frequency, F (2, 9) = 2.18, MSE = 1226176.1, p = 0.169; word length, F (2, 9) = 1.50, MSE = 0.389, p = 0.274; and orthographic neighbourhood size values, F (2, 9) = 0.525, MSE = 10.64, p = 0.609. Additionally none of the characteristics significantly correlated with each other: word frequency and word length, r (12) = -0.325, p = 0.302, word frequency and orthographic neighbourhood size, r (12) = 0.133, p = 0.680, and word length and orthographic neighbourhood size, r (12) = -0.552, p = 0.063.

HVLT total recall score. The HVLT is usually scored using a total recall score. This is the sum of all words correctly recalled, including any re-remembered words, across the three immediate trials. The total recall score was 14.9 words (SD = 5.65) for the control cases and 6.6 words (SD = 5.39) for the dementia cases. This difference was statistically significant, t (122) = 8.36, p < 0.001, with a large effect size (Cohen's d = 1.5).

Overall proportion of words recalled. Figure 1 shows the overall proportion of words recalled for the dementia and control groups for the three immediate trials of the HVLT. It is evident that recall for all trials is substantially higher for the control group compared to dementia cases. There is a greater increase in recall occurring across trials for controls (14%) but a smaller increase in recall for dementia cases (3%).

--Figure 1 about here--

Table 2 summarises a 2 (group: dementia or control) x 3 (trial: 1-3) mixed ANOVA performed to examine overall recall performance. There was a significant main effect of group; controls had significantly higher recall than dementia cases. There was a significant main effect of trial; pairwise comparisons using a Bonferroni correction indicated that recall significantly increased between each trial (trial 1 < trial 2 < trial 3). Finally, there was a significant interaction. Simple effects using a Bonferroni correction ($p = 0.00556$) comparing control and dementia cases within each trial were all significantly different (all p's < 0.001). Looking within the control cases only across trials, all comparisons were significantly different (all p's < 0.001). However, looking within the dementia cases only across trials, there were no significant comparisons (all p's > 0.100). Therefore the interaction was driven by the significantly increasing performance of controls across trials 1-3, compared to the consistent and poor performance of dementia cases across trials 1-3.

--Table 2 about here--

Category recall and accessibility. Figure 2 shows the proportion of words recalled for each category for the dementia and control groups for each of the three trials (Panel A – trial 1, Panel B – trial 2, and Panel C – trial 3). Panel D shows the overall proportion of words recalled from each category at any point during the three trials. In all four panels it is evident that recall for all categories is higher for the control group compared to dementia cases. Both control and dementia cases recalled the animal category the best, and the shelter category the worst.

--Figure 2 about here--

Table 2 summarises a series of four 2 (group: dementia or control) x 3 (category: animals, stones, and shelter) mixed ANOVAs performed to examine category recall performance in each of the four panels of Figure 2. For all ANOVAs the main effect of group was significant; controls had significantly higher recall of the categories than dementia cases. For all ANOVAs the main effect of category was significant. Bonferroni corrected pairwise comparisons indicated that for trial 2, trial 3 and overall trials 1-3 analyses all categories were recalled significantly differently (all p's < 0.001; shelter < stones < animals). However, for trial 1 animals were recalled significantly better than stones and shelter (both p's < 0.001), but stones and shelter were not different (p = 0.100). The interaction was not significant for trial 2, trial 3 and overall trials 1-3 analyses. However, the interaction was significant for trial 1. Simple effects using a Bonferroni correction (p = 0.00556) comparing control and dementia cases within each category were all significantly different (all p's < 0.004). Looking across categories within the control cases only, animals were recalled significantly better than both stones and shelter (both p's < 0.001) but there was a non-significant difference between stones and shelter (p = 0.036). Looking across

19

categories within the dementia cases only, animals were recalled significantly better than both stones and shelter (both p's < 0.002) but there was a non-significant difference between stones and shelter ($p = 0.397$). Therefore the interaction for trial 1 was driven by the difference in recall of stones compared to shelter between control and dementia cases – recall of these two categories was deemed non-significant for controls due to the Bonferroni correction but was highly non-significant for dementia cases.

The following analysis looks at category accessibility and subsequent number of words recalled to determine whether the overall category recall findings reflect different accessibility abilities for control and dementia cases. Figure 3 shows the percentage of participants who accessed each category within the three trials giving a measure of category accessibility (Panel A = trial 1, Panel B = trial 2, and Panel C = trial 3) and displays the average number of words recalled given that the category was accessed (Panel D = trial 1, Panel E = trial 2, and Panel F = trial 3). It appears that for all categories across all trials controls have a greater ability to access each category compared to controls. However, when looking at the subsequent number of words recalled there is a more consistent approximate 0.5 word advantage for control over dementia cases. It also appears that in trial 1 there is a smaller difference between control and dementia cases in both the stones and shelter categories.

--Figure 3 about here--

Table 3 summarises a series of nine Chi-square analyses performed to examine category accessibility for control and dementia cases for each category within each trial (Figure 3, Panels A-C). All Chi-squares were significant indicating that for all categories across all trials the dementia cases had impaired accessibility compared to controls.

Table 3 also summarises a series of nine independent samples t-tests performed to compare the number of words recalled given the category was accessed for control and dementia cases for each category and each trial (Figure 3, Panels D-F). Note that mixed ANOVAs were not performed due to the low number of participants who accessed every category within a trial. Controls recalled significantly more words than dementia cases for the majority of categories and trials, apart from trial 1 stones and shelter.

--Table 3 about here--

Relationship of word characteristics and recall. A series of six correlations were performed to determine whether control and dementia cases recall performance (the average recall over trials 1-3 of the HVLT) was related to the characteristics of word frequency, word length or orthographic neighbourhood size. None of the correlations were significant indicating that there is no relationship between these characteristics and recall performance. Control group: word frequency, r (12) 0.131, p = 0.685, word length, r (12) 0.065, p = 0.841, and orthographic neighbourhood size, r (12) 0.028, p = 0.932; Dementia group: word frequency, r (12) 0.256, p = 0.422, word length, r (12) -0.041, p = 0.900, and orthographic neighbourhood size, r (12) 0.197, p = 0.539.

For each characteristic within the Indonesian version the 12 words were median split into high and low frequency, short and long words, and small and large orthographic neighbourhood size[9]. Figure 4 shows the values for high and low word frequency (Panel A) and large and small orthographic neighbourhood size (Panel B) for both dementia and control groups.

---

[9] Word length was not analysed further due to the unequal median split (low = 9 words, high = 3 words)

21

--Figure 4 about here--

Table 4 summarises two 2 (group: dementia or control) x 2 (word characteristic: low or high) x 3 (trial: 1-3) mixed ANOVAs performed separately on the word frequency and orthographic neighbourhood size word characteristics to examine whether they have an effect on recall. The *overall proportion of words recalled* analysis reported earlier summarises the findings for the main effects of group and trial and the two-way interaction between them. Within the current analysis we are interested in the effects of word frequency and orthographic neighbourhood size. For word frequency the main effect and two-way interaction with trial were significant; high frequency words were recalled better than low frequency words. Simple effects using a Bonferroni correction (p = 0.00556) comparing high and low frequency within each trial indicated that low frequency words were remembered significantly worse than high frequency words for trial 1 (p < 0.001) but there were no significant differences for trials 2 and 3 (both p's < 0.018). Looking across trials within the high frequency words only, recall stayed the same across trials (trial 1 = trial 2 = trial 3). Looking across trials within the low frequency words only, recall significantly increased between each trial (trial 1 < trial 2 < trial 3). Therefore the interaction was driven by improvement in recall for the low frequency words as the test progressed. However, the two-way interaction between word frequency and group and the three way interaction between word frequency, group and trial were not significant. For orthographic neighbourhood size the main effect was significant; words with fewer orthographic neighbours were recalled better than words with more orthographic neighbours. However, all interactions were non-significant. Therefore both word frequency and orthographic neighbourhood size did not affect dementia cases and controls differently.

--Table 4 about here--

Serial position curves. Figure 5 shows the serial position curves for the dementia and control groups for each of the three trials (Panel A – trial 1, Panel B – trial 2, and Panel C – trial 3). Panel D shows how likely each word was to be recalled across the three trials. Recall in all trials is higher for the control group compared to dementia cases. Both controls and dementia cases show increased recall for the start and end of the list, indicating primacy and recency respectively. However, for dementia cases in particular, this may well be a consequence of the placement of the words; serial positions 1, 3 and 11 are words from the highly recalled animal category.

--Figure 5 about here--

Table 5 summarises a series of four 2 (group: dementia or control) x 12 (serial position: 1-12) mixed ANOVAs performed to examine serial position performance in each of the four panels of Figure 5. For all ANOVAs the main effect of group was significant; controls had significantly higher recall than dementia cases. In addition, the main effect of serial position was significant; Bonferroni corrected pairwise comparisons indicated primacy and recency (the exact statistics can be found in Appendix 2 for trial 1, Appendix 3 for trial 2, Appendix 4 for trial 3, and Appendix 5 for trials 1-3). The interaction was not significant for trial 2, trial 3 and overall trials 1-3 analyses. However, the interaction was significant for trial 1. Simple effects compared performance between the groups at each serial position only, as this was the main comparison of interest. Using a Bonferroni correction ($p = 0.00417$), recall

for dementia cases was significantly lower than controls at serial positions 2, 3, and 9 (all p's < 0.002). Therefore the interaction was driven by similar recall for dementia cases and controls except for serial positions 2, 3, and 9.

--Table 5 about here--

Retrieval dynamics across trials. Figures 2 and 5 hint at the learning and forgetting that occurred across the trials. Within Figure 5 the dementia cases appeared to demonstrate low levels of learning across trials (i.e. adding new words), whereas the control group demonstrated higher levels particularly for the asymptote (middle) serial positions. However, Figure 5 only indicates overall increases or decreases in specific words over trials. It does not allow for a detailed understanding of the learning, retention and forgetting across trials for individual participants. Therefore for each participant, for each word, and across all trials, the number of times they retained, learnt or forgot a word was calculated in order to assess retrieval dynamics.

Table 6 shows the proportion of learning, retention and forgetting occurring across trials 2 and 3 for control and dementia cases. The data was divided into four categories: (1) *Re-remembered* means that a word was recalled on the current trial and had been recalled on previous trial(s) i.e., retention, (2) *New* means that it was recalled on the current trial but had not been recalled on any previous trial(s) i.e. learning, (3) *Forgotten* means that it was not recalled on the current trial but had been recalled on previous trial(s), (4) *Never recalled* means that it was not recalled on the current trial and had not been recalled on any previous trial(s). Control cases showed much greater levels of re-remembering and lower levels of words never recalled for trials 2 and 3 compared to dementia cases. However, the rates of

new words learnt and words forgotten for trials 2 and 3 appeared to be more similar for both control and dementia cases.

--Table 6 about here--

Table 7 summarises a series of four 2 (group: dementia or control) x 2 (trial: 2-3) mixed ANOVAs performed to examine the levels of re-remembered, new, forgotten, and never recalled words for controls and dementia cases shown in Table 7. For all ANOVAs the main effect of group was significant; compared to controls, dementia cases re-remembered, acquired, and forgot fewer words and never recalled more words. For all ANOVAs the main effect of trial was significant; fewer words were re-remembered and forgotten, and more words were acquired and never recalled in trial 2 compared to trial 3. The interaction was not significant for the new or forgotten words analyses. However, the interactions were significant for the re-remembered and never recalled words analyses. Looking first at the re-remembered interaction simple effects, using a Bonferroni correction ($p = 0.0125$), indicated that controls re-remembered significantly more than dementia cases for both trial 2 and 3 (both p's < 0.001) and the amount of re-remembered words significantly increased between trial 2 and 3 for both control and dementia cases (both p's < 0.002). Therefore, this interaction was driven by dementia cases showing significantly lower levels of re-remembering than controls for both trials 2 and 3. Looking now at the never recalled interaction simple effects, using a Bonferroni correction ($p = 0.0125$), indicated that the amount of words never recalled was significantly higher for dementia cases compared to controls for both trial 2 and 3 (both p's < 0.001) and the amount of words never recalled significantly decreased between trial 2 and 3 for both control and dementia cases (both p's $<$

0.001). Therefore, the interaction was driven by dementia cases showing significantly higher levels of words never recalled than controls for both trials 2 and 3.

--Table 7 about here--

Due to the difference between the number of words of the HVLT list remembered by controls and dementia cases, perhaps a more accurate comparison to assess retrieval dynamics would be to calculate the proportion of words acquired and forgotten *relative* to the proportion of the list previously recalled. For example, dementia cases appear to have forgotten more words than controls in the above analysis but this is because they remember less than controls to start with. Looking at the words forgotten relative to the proportion of the list previously recalled, controls show the same proportion of words that they forgot across the two trials (trial 2 = 29% of the list previously remembered, trial 3 = 29% of the list previously remembered). Dementia cases also show a very similar, albeit higher than controls, proportion of words that they forgot across the two trials (trial 2 = 41% of the list previously remembered, trial 3 = 42% of the list previously remembered). Looking at the words newly acquired relative to the proportion of the list previously recalled, controls show a very similar proportion of newly acquired words across the two trials (trial 2 = 26% of the list yet to be recalled, trial 3 = 27% of the list yet to be recalled). However, dementia cases show a lower- but greater- difference in the proportion of the remaining list that they acquire across the two trials (trial 2 = 11% of the list yet to be recalled, trial 3 = 7% of the list yet to be recalled).

Discussion

The aim of this paper was to investigate in-depth recall patterns within the HVLT to determine whether there are important subtle differences between control and dementia cases

recall that are not currently being captured using the total immediate recall score. We hoped that any novel findings might lead to improvement in the efficacy of the HVLT.

As expected the dementia cases had a much lower total immediate recall score on the HVLT compared to controls. However, we found dementia cases exhibited widespread impairments across all analyses. The dementia cases showed no improvement in recall across the three trials in agreement with previous research of diminished learning rates (Delis et al., 1991; Ober et al., 1985; Perri et al., 2005; Weingartner et al., 1981, 1983). Focusing on recall within the three categories of the HVLT, dementia cases showed poorest recall for the shelter category and best recall for the animals category; although for trial 1 recall of stones and shelter was not different. These trends were mirrored in the control cases, albeit at higher recall values. This is in contrast to Schrijnemaekers et al. (2006) who found no significant difference between recall of the three categories. Although their trends revealed animals were recalled best, stones and shelter were equally poorly recalled. However, it is important to note that Schrijnemaekers et al. study utilised an English sample and the English version of the HVLT whereas we utilised an Indonesian sample and an Indonesian version of the HVLT. Cultural differences can impact on category recall, possibly related to the following linguistic aspects.

The linguistic characteristics we were able to analyse did not reveal significant differences on any of our measures between the categories. Therefore another linguistic feature that we could not measure may be driving these differences. Age-of-acquisition is a potential characteristic that could be causing these differences. The English age-of-acquisition values showed that animals were acquired significantly earlier than both stones and shelter with Schrijnemaekers et al. (2006) non-significant findings exactly mirroring these trends. Dementia cases show a robust advantage for early acquired words over a range of tasks (Cuetos et al., 2010; Forbes-McKay et al., 2005; Kremin et al., 2000; Ralph et al.,

1998; Silveri et al., 2002). We also found dementia cases had reduced accessibility to all categories within all trials compared to controls. Given that a category was accessed, dementia cases subsequent recall was also impaired relative to controls in accord with previous research (Martin & Fedio, 1983; Ober, Dronkers, Koss, Delis & Friedland, 1986).

Dementia cases often display a reduced primacy effect, but a relatively preserved recency effect relative to controls (e.g., Bayley et al., 2000; Carlesimo et al., 1996; Foldi et al., 2003; Gainotti & Marra, 1994; Howieson et al., 2011; Massman et al., 1993; Tierney et al., 1994). However, we found reduced levels of both primacy and recency for dementia cases compared to controls. This discrepant finding may be due to the recency effect declining with the severity of dementia (Pepin & Eslinger, 1989). Whilst these levels were reduced, dementia cases did show significant levels of primacy and recency within their recall. However, we are concerned that the higher proportion of the well-remembered animals category at the start and end of the list (serial positions 1, 3 and 11) may be inflating primacy and recency levels. Due to this confound, it is unclear whether the improved performance is due to one or both of these factors. In addition three of the words from the shelter category are in the middle serial positions (4, 6 and 7) which may contribute to their poorer recall. Other memory tests used to screen for dementia present the words in a random order each time (e.g., CVLT, Delis et al., 1987; CERAD word learning list, Welsh et al., 1991) which eliminates these confounds. However, this may be one of the reasons why floor effects are not usually reported on the HVLT. To help disentangle these issues future work could manipulate the presentation order of the words by either presenting the words in a random order for each trial or, perhaps more interestingly, by systematically varying the location of the categories within the list such that that one word of each category occurs in all four sections of the list.

Looking in detail at the intertrial retrieval dynamics revealed that dementia cases did recall new words in trials 2 and 3 despite not showing a significant increase in learning from trial 1 to 3 overall. However, dementia cases did learn significantly less than controls. Dementia cases also recall significantly less of the list and, relative to the proportion of the list already recalled, also forgot more words, mirroring the findings for gained and lost access (Genon et al., 2013; Moulin et al., 2004; Woodward et al., 1999). Dementia cases appear to have deficits in both learning and consolidation indicating issues with encoding and storage, respectively. It would appear from these results and previous findings (Algarabel et al., 2009, 2012; Ally et al., 2009; Butters et al., 1983, 1987; Carlesimo et al., 1998; Dallabarba & Wong, 1995; Dannenbaum et al., 1988; Delis et al., 1991; Gaines et al., 2006; Genon et al., 2013; Glosser et al., 2002; Granholm & Butters, 1988; Grober & Buschke, 1987; Hodges et al., 1999; Koen & Yonelinas, 2014; Kopelman, 1991; Martin et al., 1985; Moulin et al., 2004; Ober et al., 1985; Perri et al. 2005; Weingartner et al., 1981, 1983; Wolk et al., 2011; Woodward et al., 1999) that dementia cases have impairment at all stages of memory (encoding, storage, and retrieval).

What was interesting were the relative rates across trials 2 and 3 of the proportion of newly acquired words and the loss of previously recalled words. With regards to immediate forgetting rates, the proportion of forgotten words changed by less than 1% across trials 2 and 3 for both control and dementia cases, showing that whilst immediate forgetting rates are higher in dementia cases, they follow a similar trend. Acquisition rates are slightly more divergent between dementia cases and controls. Controls show a stable rate of acquisition across trials 2 and 3, however, dementia cases show a decrease in acquisition rates, even though in absolute terms they have a much larger number of never recalled words to draw from. These rates of immediate forgetting and, in particular, acquisition might be a useful diagnostic tool alongside the total recall score to be able to differentiate between dementia

29

and other memory disorders, such as aMCI, which often displays results in-between dementia cases and controls.

Additionally we also looked at several linguistic characteristics of the words within the HVLT. This data set was collected with Indonesian speaking participants but we analysed the linguistic features of the 12 words of Form 1 of the HVLT for both English and Indonesian languages as these characteristics have never been researched before, in either language or within the HVLT. We found that the two languages provided broadly similar findings. The English and Indonesian languages were strongly correlated for both word frequency and orthographic neighbourhood size characteristics. Within each language, all but one characteristic did not correlate with each other, indicating that these linguistic measures are relatively independent. For both languages word frequency, word length, and orthographic neighbourhood size did not vary across the three categories of Form 1 of the HVLT. However, there were two differences between the two languages. Firstly, there was no correlation between the English and Indonesian languages for word length, although this is most likely due to the lack of variation within the data. Secondly, within the English language there was a relationship between word length and orthographic neighbourhood size which was non-significant within the Indonesian language. Therefore the use of Indonesian language may have partially affected the results we may have expected with an English speaking sample. We also found age-of-acquisition did vary across the categories within the English version, but no comparison with Indonesian was possible due to the lack of values for age-of-acquisition. Therefore, we were also unable to determine the possible effect of age-of-acquisition on recall within the current data. Previous research has not looked at recall (Cuetos et al., 2010; Forbes-McKay et al., 2005; Kremin et al., 2000; Ralph et al., 1998; Silveri et al., 2002). We would be keen to address this within future research.

We also investigated the effect of these characteristics on recall. We found that neither word frequency nor orthographic neighbourhood size had a differential effect on recall in control and dementia cases. However, we did find an overall word frequency effect; high frequency words were recalled better than low frequency words, for both control and dementia cases. This is in line with previous findings from recall tasks for controls (Deese, 1960; Postman, 1970; Sumby, 1963) and the only study to investigate dementia cases with free recall (Hessler et al., 2016). We also found an overall effect of orthographic neighbourhood size; where words with fewer orthographic neighbours produced better recall for both control and dementia cases. Hessler et al. (2016) also found better recall with fewer orthographic neighbours for dementia cases. However, they found the opposite pattern in control cases. Previous research using serial recall also supports Hessler et al.'s finding for controls (Allen & Hulme, 2006; Glanc & Greene, 2012; Jalbert, Neath, Bireta & Surprenant, 2011; Jalbert, Neath & Surprenant, 2011; Roodenrys et al., 2002). We are not certain why our results differ, but one possibility may be due to the much smaller vocabulary of Indonesian compared to German and English affording the possibility of fewer orthographic neighbours. Further research investigating the effect of orthographic neighbourhood size within free recall would address this current discrepancy.

These linguistic findings reveal the importance of the appropriate and careful selection of words for verbal recall tests and have an impact on the construction of future word based dementia screening tests and the potential updating of current tests. Much research in dementia uses the levels of primacy and recency as a marker for diagnosis. However, the tests that they utilise do not look at the linguistic characteristics of each individual word. The effects of these four characteristics may not be uniform across the list and therefore be producing a confound between serial position and these characteristics. For example, word frequency appears to affect pre-recency locations (Raymond, 1969; Sumby,

31

1963). However, more recently word frequency has been found to affect recency positions (Van Overschelde, 2002) but these discordant findings may be due to the use of pure and mixed frequency lists. Word frequency may also affect where recall is initiated within the list which then has a knock-on effect in recall. Recall is more likely to be initiated with the first word in the list for high frequency words, but towards the end of the list for low frequency words (Sumby, 1963). Grenfell-Essam and colleagues have shown that where recall is initiated has a large effect on primacy and recency (Grenfell-Essam & Ward, 2012, 2015; Grenfell-Essam et al., 2013, 2017; Ward et al., 2010).

We have shown that both word frequency and orthographic neighbourhood size affect recall, with the effect of age-of-acquisition unknown. As the HVLT repeats the words in the same order in each list this makes an in-depth analyses of the linguistic characteristics of the five other forms of the HVLT all the more important. In the future to avoid either floor or ceiling effects, the difficulty of tests can be altered by adjusting the word frequency, word length, age-of-acquisition and/or orthographic neighbourhood size. Indeed the CVLT-II utilises higher frequency words specifically to avoid floor effects (Delis, Kramer, Kaplan & Ober, 2000). Researchers need to also carefully consider the location of these items within the list and whether the words are presented in the same, or random, order.

This study does have some limitations. Our algorithm for dementia was validated in several studies in Indonesia (see Hogervorst et al., 2011), However, future work needs to further validate this work in clinically established dementia cases and controls. Another limitation is the translation of the test which does not allow transference of results to other cultures. Future work should look at a range of cultures to investigate the generalisability of our results. The translation of this test into Indonesian also did not allow for the characteristic of age-of-acquisition to be assessed or use of established word frequency and orthographic

neighbourhood size corpora. We see this research as the first step in developing a validated version of the HVLT.

There are several ways in which this research can be pursued in the future. Firstly, the fact that the HVLT has six different forms is a strength of the test, as it allows for repeated testing (Benedict et al., 1998; Krebs, 1994; Rasmusson et al., 1995; Woods et al., 2005). We wish to collect data that would allow us to extend our novel set of analyses to determine whether the six forms are actually equivalent based on the findings from Form 1 in this paper. Secondly, it was unknown at the time of data collection that the data would be used for in-depth analysis and therefore output order was not recorded. Output order data would have enabled recall consistency and semantic clustering analyses to be performed that are known to be reduced in dementia recall (e.g., Carlesimo et al., 1998; Delis et al., 1991; Gaines et al., 2006; Glosser et al., 2002; Hodges et al., 1999; Perri et al., 2005; Weingartner et al., 1981). Recording output order would also have allowed for analysis of where in the list participants initiate their recall. There may be differences in the starting point of recall between controls and dementia patients, and also differences in whether the same word is chosen as the first word in trials 1, 2 and 3. Output order also allows the full sequence to be investigated allowing for assessment, for each participant, of how similar recall is from trial to trial. Thirdly, we would also like to reproduce this work, along with the added output order analyses, in an English sample as changes to some of the stones were necessary to accommodate local knowledge and we were also unable to examine the possible effect of age-of-acquisition. The linguistic characteristics may have a large effect on primacy and recency. Finally, we would also be keen to extend this in-depth analysis of HVLT with other populations, such as vascular dementia and very early dementia cases (e.g. possible Alzheimer's disease or aMCI), to see whether there are distinct patterns of deficits that can be identified to once again help improve the efficacy of this test. Indeed, due to the widespread

impairments of dementia cases across all performance measures in this study, it might be more useful to profile earlier deficits in patients with aMCI. aMCI patients may show less widespread patterns of impairment that when tracked over time might be shown to be more likely to lead to dementia.

In conclusion, this paper has demonstrated that there are important differences between control and dementia case recall that are not captured in the total recall score. We have demonstrated that, within the HVLT, dementia cases exhibit: consistent and poor recall across the three trials, poorest recall for shelter, but best for animals, impaired accessibility of all categories and reduced subsequent recall from accessed categories, reduced recall at the start and end of the list, lower levels of re-remembering and recalling new words; and higher levels of forgetting and never recalling words. These differences help characterise in more detail the exact nature of the reduced recall performance of dementia cases compared to controls, and when used alongside the total recall score, may help improve the efficacy of the HVLT. This research may help improve the diagnostic ability and accuracy of the HVLT to detect earlier stages of dementia than currently possible. It could also be valuable when monitoring individuals over time to identify subtle changes in HVLT performance not captured with the total recall score. Finally, it might prove beneficial in assessment of treatment efficacy. Therefore, going forward this research may have wide reaching implications.

References

Algarabel, S., Escudero, J., Mazón, J. F., Pitarque, A., Fuentes, M., Peset, V., & Lacruz, L. (2009). Familiarity-based recognition in the young, healthy elderly, mild cognitive impaired and Alzheimer's patients. *Neuropsychologia, 47(10),* 2056-2064.

Algarabel, S., Fuentes, M., Escudero, J., Pitarque, A., Peset, V., Mazón, J. F., & Meléndez, J. C. (2012). Recognition memory deficits in mild cognitive impairment. *Aging, Neuropsychology, and Cognition, 19(5)*, 608-619.

Allen, R., & Hulme, C. (2006). Speech and language processing mechanisms in verbal serial recall. *Journal of Memory and Language, 55*, 64-88.

Ally, B. A., Gold, C. A., & Budson, A. E. (2009). An evaluation of recollection and familiarity in Alzheimer's disease and mild cognitive impairment using receiver operating characteristics. *Brain and cognition, 69(3),* 504-513.

Balota, D. A., Burgess, G. C., Cortese, M. J., & Adams, D. R. (2002). The word-frequency mirror effect in young, old, and early-stage Alzheimer's disease: Evidence for two processes in episodic recognition performance. *Journal of Memory and Language, 46(1),* 199-226.

Barry, C., Morrison, C. M., & Ellis, A. W. (1997). Naming the Snodgrass and Vanderwart pictures: Effects of age of acquisition, frequency and name agreement. *Quarterly Journal of Experimental Psychology, 50(A),* 560 –585

Bayley, P. J., Salmon, D. P., Bondi, M. W., Bui, B. K., Olichney, J., Delis, D. C., Thomas, R. G., & Thal, L. J. (2000). Comparison of the serial position effect in very mild Alzheimer's disease, mild Alzheimer's disease, and amnesia associated with electroconvulsive therapy. *Journal of the International Neuropsychological Society, 6(03),* 290-298.

Belleville, S., Peretz, I., & Malenfant, D. (1996). Examination of the working memory components in normal aging and in dementia of the Alzheimer type. *Neuropsychologia, 34(3),* 195-207.

Grenfell-Essam, R., Hogervorst, E. & Rahardjo, T. B. W. (in prep)

Benedict, R. H., Schretlen, D., Groninger, L., & Brandt, J. (1998). Hopkins Verbal Learning Test–Revised: Normative data and analysis of inter-form and test-retest reliability. *The Clinical Neuropsychologist, 12(1),* 43-55.

Benedict, R. H., Schretlen, D., Groninger, L., Dobraski, M., & Shpritz, B. (1996). Revision of the Brief Visuospatial Memory Test: Studies of normal performance, reliability, and validity. *Psychological Assessment, 8(2),* 145-153.

Brandt, J. (1991). The Hopkins Verbal Learning Test: Development of a new memory test with six equivalent forms. *Clinical Neuropsychologist*, 5(2), 125-142.

Brysbaert, M., & New, B. (2009). Moving beyond Kucera and Francis: A Critical Evaluation of Current Word Frequency Norms and the Introduction of a New and Improved Word Frequency Measure for American English. *Behavior Research Methods, 41(4),* 977-990.

Butters, N., Albert, M. S., Sax, D. S., Miliotis, P., Nagode, J., & Sterste, A. (1983). The effect of verbal mediators on the pictorial memory of brain-damaged patients. *Neuropsychologia, 21(4),* 307-323.

Butters, N., Granholm, E., Salmon, D. P., Grant, I., & Wolfe, J. (1987). Episodic and semantic memory: A comparison of amnesic and demented patients. *Journal of Clinical and Experimental Neuropsychology, 9(5),* 479-497.

Carlesimo, G. A., Mauri, M., Graceffa, A. M., Fadda, L., Loasses, A., Lorusso, S., & Caltagirone, C. (1998). Memory performances in young, elderly, and very old healthy individuals versus patients with Alzheimer's disease: Evidence for discontinuity between normal and pathological aging. *Journal of clinical and experimental neuropsychology, 20(1),* 14-29.

Carlesimo, G. A., Fadda, L., Sabbadini, M., & Caltagirone, C. (1996). Recency effect in
Alzheimer's disease: a reappraisal. *Quarterly Journal of Experimental Psychology A,
Human Experimental Psychology, 49*, 315-325.

Carroll, J. B., & White, M. N. (1973). Word frequency and age-of acquisition as determiners
of picture-naming latency. *Quarterly Journal of Experimental Psychology, 25,* 85–95.

Cherner, M., Suarez, P., Lazzaretto, D., Artiola i Fortuny, L., Mindt, M. R., Dawes, S.,
Marcotte, T., Grant, I., Heaton, R., & HNRC group (2007). Demographically
corrected norms for the Brief Visuospatial Memory Test-revised and Hopkins Verbal
Learning Test-revised in monolingual Spanish speakers from the US–Mexico border
region. *Archives of Clinical Neuropsychology, 22(3),* 343-353.

Christian, J., Bickley, W., Tarka, M., & Clayton, K. (1978). Measures of free recall of 900
English nouns: Correlations with imagery, concreteness, meaningfulness, and
frequency. *Memory & Cognition, 6(4),* 379-390.

Coltheart, M., Davelaar, E., Jonasson, J. T., & Besner, D. (1977). Access to the internal
lexicon. In S. Dornic (Ed.), *Attention and performance VI* (pp. 535-555). Hillsdale,
NJ: Erlbaum.

Coltheart, V., & Winograd, E. (1986). Word imagery but not age of acquisition affects
episodic memory. *Memory & Cognition, 14(2),* 174-180.

Cuetos, F., Herrera, E., & Ellis, A. W. (2010). Impaired word recognition in Alzheimer's
disease: The role of age of acquisition. *Neuropsychologia, 48(11),* 3329-3334.

Dallabarba, G., & Wong, C. (1995). Encoding specificity and intrusion in Alzheimers-disease
and amnesia. *Brain and cognition, 27(1),* 1-16.

Dannenbaum, S. E., Parkinson, S. R., & Inman, V. W. (1988). Short-term forgetting:
Comparisons between patients with dementia of the Alzheimer type, depressed, and
normal elderly. *Cognitive Neuropsychology, 5(2),* 213-233.

Deese, J. (1960). Frequency of usage and number of words in free recall: The role of association. *Psychological Reports, 7*, 337-344.

de Jager, C. A., Schrijnemaekers, A. M. C., Honey, T. E. M., & Budge, M. M. (2009). Detection of MCI in the clinic: evaluation of the sensitivity and specificity of a computerised test battery, the Hopkins Verbal Learning Test and the MMSE. *Age and ageing, 1-6.*

Delis, D. C., Kramer, J. H., Kaplan, E., & Ober, B. A. (2000). *CVLT-II: California verbal learning test: adult version.* Psychological Corporation.

Delis, D. C., Kramer, J. H., Kaplan, E., & Thompkins, B. A. O. (1987). *CVLT: California Verbal Learning Test-Adult Version: Manual*. Psychological Corporation.

Delis, D. C., Massman, P. J., Butters, N., Salmon, D. P., Cermak, L. S., & Kramer, J. H. (1991). Profiles of demented and amnesic patients on the California Verbal Learning Test: Implications for the assessment of memory disorders. *Psychological Assessment: A Journal of Consulting and Clinical Psychology, 3(1),* 19-26.

Dewhurst, S. A., Hitch, G. J., & Barry, C. (1998). Separate effects of word frequency and age of acquisition in recognition and recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24(2)*, 284-298.

Dunabeitia, J. A., Marín, A., & Carreiras, M. (2009). Associative and orthographic neighborhood density effects in normal aging and Alzheimer's disease. *Neuropsychology, 23(6),* 759-764.

Foldi, N. S., Brickman, A. M., Schaefer, L. A., & Knutelska, M. E. (2003). Distinct serial position profiles and neuropsychological measures differentiate late life depression from normal aging and Alzheimer's disease. *Psychiatry research, 120(1),* 71-84.

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). "Mini-mental state": a practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, 12(3), 189-198.

Forbes-McKay, K. E., Ellis, A. W., Shanks, M. F., & Venneri, A. (2005). The age of acquisition of words produced in a semantic fluency task can reliably differentiate normal from pathological age related cognitive decline. *Neuropsychologia, 43(11),* 1625-1632.

Frank, R. M., & Byrne, G. J. (2000). The clinical utility of the Hopkins Verbal Learning Test as a screening test for mild dementia. *International journal of geriatric psychiatry, 15(4),* 317-324.

Friedman, M. A., Schinka, J. A., Mortimer, J. A., & Graves, A. B. (2002). Hopkins verbal learning test–revised: Norms for elderly African Americans. *The Clinical Neuropsychologist, 16(3),* 356-372.

Gaines, J. J., Shapiro, A., Alt, M., & Benedict, R. H. B. (2006). Semantic Clustering Indexes for the Hopkins Verbal Learning Test Revised: Initial Exploration in Elder Control and Dementia Groups. *Applied Neuropsychology, 13(4),* 213-222.

Gainotti, G., & Marra, C. (1994). Some aspects of memory disorders clearly distinguish dementia of the Alzheimer's type from depressive pseudo-dementia. *Journal of Clinical and Experimental Neuropsychology, 16(1),* 65-78.

Ganguli, M., Ratcliff, G., Chandra, V., Sharma, S., Gilby, J., Pandav, R., Belle, S., Ryan, C., Baker, C., Seaberg, E., & Dekosky, S. (1995). A Hindi version of the MMSE: the development of a cognitive screening instrument for a largely illiterate rural elderly population in India. *International Journal of Geriatric Psychiatry*, 10(5), 367-377.

Genon, S., Collette, F., Moulin, C. J., Lekeu, F., Bahri, M. A., Salmon, E., & Bastin, C. (2013). Verbal learning in Alzheimer's disease and mild cognitive impairment: fine-

Grenfell-Essam, R., Hogervorst, E. & Rahardjo, T. B. W. (in prep)

grained acquisition and short-delay consolidation performance and neural correlates. *Neurobiology of aging, 34(2),* 361-373.

Gerhand, S., & Barry, C. (1998). Word frequency effects in oral reading are not merely age-of-acquisition effects in disguise. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24,* 267–283.

Gilhooly, K. J., & Gilhooly, M. L. (1979). Age-of-acquisition effects in lexical and episodic memory tasks. *Memory & Cognition, 7(3),* 214-223.

Glanc, G., & Greene, R. (2012). Orthographic distinctiveness and memory for order. *Memory, 20(8),* 865-871.

Glosser, G., Gallo, J. L., Clark, C. M., & Grossman, M. (2002). Memory encoding and retrieval in frontotemporal dementia and Alzheimer's disease. Neuropsychology, 16(2), 190-196.

Granholm, E., & Butters, N. (1988). Associative encoding and retrieval in Alzheimer's and Huntington's disease. *Brain and cognition, 7(3),* 335-347.

Grenfell-Essam, R., & Ward, G. (2012). Examining the relationship between free recall and immediate serial recall: The role of list length, strategy use, and test expectancy. *Journal of Memory and Language, 67,* 106-148.

Grenfell-Essam, R. & Ward, G. (2015). The effect of selective attention and a stimulus prefix on the output order of immediate free recall of short and long lists. *Canadian Journal of Experimental Psychology, 69(1),* 1-16.

Grenfell-Essam, R., Ward, G., & Tan, L. (2013). The role of rehearsal on the output order of immediate free recall of short and long lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39,* 317-347

Grenfell-Essam, R., Ward, G., & Tan, L. (2017). Common Modality Effects in Immediate Free Recall and Immediate Serial Recall. *Journal of Experimental Psychology:*

Learning, Memory, and Cognition. Advance online publication.

http://dx.doi.org/10.1037/xlm0000430.

Grober, E., & Buschke, H. (1987). Genuine memory deficits in dementia. *Developmental neuropsychology, 3(1),* 13-36.

Hermit Dave (2016). 2016 *Frequency list files generated from http://opus.lingfil.uu.se/OpenSubtitles2016.php*. Retrieved from https://github.com/hermitdave/FrequencyWords/blob/master/content/2016/id/id_full.t xt

Hessler, J. B., Fischer, A. M., & Jahn, T. (2016). Differential Linguistic Recall Effects in the California Verbal Learning Test in Healthy Aging and Alzheimer's Dementia: Analysis of Routine Diagnostic Data. *Archives of Clinical Neuropsychology, 31(7),* 689-699.

Hester, R. L., Kinsella, G. J., Ong, B., & Turner, M. (2004). Hopkins verbal learning test: Normative data for older Australian adults. *Australian Psychologist, 39(3),* 251-255.

Hodges, J. R., Patterson, K., Ward, R., Garrard, P., Bak, T., Perry, R., & Gregory, C. (1999). The differentiation of semantic dementia and frontal lobe dementia (temporal and frontal variants of frontotemporal dementia) from early Alzheimer's disease: a comparative neuropsychological study. *Neuropsychology, 13(1),* 31-40.

Hogervorst, E., Combrinck, M., Lapuerta, P., Rue, J., Swales, K., & Budge, M. (2002). The Hopkins verbal learning test and screening for dementia. *Dementia and geriatric cognitive disorders, 13(1),* 13-20.

Hogervorst, E., Mursjid, F., Ismail, R.I., Prasetyo, S., Nasrun, M., Mochtar, T.N., Bandelow, S., Subarkah, L.K., & Rahardjo, T.B.W. (2011). Validation of two short dementia screening tests in Indonesia. In S. R. Jacobsen (Eds.), *Vascular dementia: risk factors, diagnosis and treatment*. New York: Nova Sci Publ, Inc.

Grenfell-Essam, R., Hogervorst, E. & Rahardjo, T. B. W. (in prep)

Howieson, D. B., Mattek, N., Seeyle, A. M., Dodge, H. H., Wasserman, D., Zitzelberger, T., & Jeffrey, K. A. (2011). Serial position effects in mild cognitive impairment. *Journal of Clinical and Experimental Neuropsychology, 33(3),* 292-299.

Jahnke, J. C. (1965). Primacy and recency effects in serial-position curves of immediate recall. *Journal of Experimental Psychology, 70*, 130-132.

Jalbert, A., Neath, I., Bireta, T. J., & Surprenant, A. M. (2011). When does length cause the word length effect? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37(2)*, 338-353.

Jalbert, A., Neath, I., & Surprenant, A. M. (2011). Does length or neighborhood size cause the word length effect? *Memory & cognition, 39(7),* 1198-1210.

Kinsbourne, M., & George, J. (1974). The mechanism of the word-frequency effect on recognition memory. *Journal of Verbal Learning and Verbal Behavior, 13*, 63-69.

Koen, J. D., & Yonelinas, A. P. (2014). The effects of healthy aging, amnestic mild cognitive impairment, and Alzheimer's disease on recollection and familiarity: a meta-analytic review. *Neuropsychology review, 24(3),* 332-354.

Kopelman, M. D. (1991). Non-verbal, short-term forgetting in the alcoholic Korsakoff syndrome and Alzheimer-type dementia. *Neuropsychologia, 29(8),* 737-747.

Krebs, R. (1994). The Hopkins Verbal Learning Test: An Alternative to the MMSE?. *The Gerontologist, 34,* 692.

Kremin, H., Hamerel, M., Dordain, M., De Wilde, M., & Perrier, D. (2000). Age of acquisition and name agreement as predictors of mean response latencies in picture naming of French adults. *Brain and Cognition, 43,* 286–291.

Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30,000 English words. *Behavior Research Methods, 44(4),* 978-990.

Kuslansky, G., Katz, M., Verghese, J., Hall, C. B., Lapuerta, P., LaRuffa, G., & Lipton, R. B. (2004). Detecting dementia with the Hopkins verbal learning test and the mini-mental state examination. *Archives of Clinical Neuropsychology, 19(1),* 89-104.

Lacritz, L. H., & Cullum, C. M. (1998). The Hopkins Verbal Learning Test and CVLT: a preliminary comparison. *Archives of Clinical Neuropsychology, 13(7),* 623-628.

Lacritz, L. H., Cullum, C. M., Weiner, M. F., & Rosenberg, R. N. (2001). Comparison of the Hopkins Verbal Learning Test-revised to the California Verbal Learning Test in Alzheimer's disease. *Applied Neuropsychology, 8(3),* 180-184.

Lawton, M. P., & Brody, E. M. (1969). Assessment of older people: self-maintaining and instrumental activities of daily living. *Gerontologist, 9(3),* 179-186.

Malek-Ahmadi, M., Raj, A., & Small, B. J. (2011). Semantic clustering as a neuropsychological predictor for amnestic-MCI. *Aging, Neuropsychology, and Cognition, 18(3),* 280-292.

Marian, V., Bartolotti, J., Chabal, S., & Shook, A. (2012). CLEARPOND: Cross-Linguistic Easy-Access Resource for Phonological and Orthographic Neighborhood Densities. *PLoS ONE 7(8)*, e43230. doi:10.1371/journal.pone.0043230.

Martin, A., Brouwers, P., Cox, C., & Fedio, P. (1985). On the nature of the verbal memory deficit in Alzheimer's disease. *Brain and Language, 25(2),* 323-341.

Martin, A., & Fedio, P. (1983). Word production and comprehension in Alzheimer's disease: The breakdown of semantic knowledge. *Brain and language, 19*, 124-141.

Massman, P. J., Delis, D. C., & Butters, N. (1993). Does impaired primacy recall equal impaired long-term storage?: Serial position effects in Huntington's disease and Alzheimer's disease. *Developmental Neuropsychology, 9(1),* 1-15.

Morris P. E. (1981). Age of acquisition, imagery, recall, and the limitations of multiple-regression analysis. *Memory & Cognition, 9(3),* 277-282.

Morris, R. G. (1984). Dementia and the functioning of the articulatory loop system. *Cognitive Neuropsychology, 1(2),* 143-157.

Morrison, C. M., & Ellis, A. W. (1995). Roles of word frequency and age of acquisition in word naming and lexical decision. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 116 –133.

Moulin, C. J., James, N., Freeman, J. E., & Jones, R. W. (2004). Deficient acquisition and consolidation: intertrial free recall performance in Alzheimer's disease and mild cognitive impairment. *Journal of Clinical and Experimental Neuropsychology, 26(1),* 1-10.

Murdock, B. B., Jr. (1962). The serial position effect of free recall. *Journal of Experimental Psychology, 64*, 482-488.

Ober, B. A., Dronkers, N. F., Koss, E., Delis, D. C., & Friedland, R. P. (1986). Retrieval from semantic memory in Alzheimer-type dementia. *Journal of clinical and experimental neuropsychology, 8(1),* 75-92.

Ober, B. A., Koss, E., Friedland, R. P., & Delis, D. C. (1985). Processes of verbal memory failure in Alzheimer-type dementia. *Brain and Cognition, 4(1),* 90-103.

Pepin, E. P., & Eslinger, P. J. (1989). Verbal memory decline in Alzheimer's disease A multiple-processes deficit. *Neurology, 39(11),* 1477-1477.

Perri, R., Carlesimo, G. A., Serra, L., Caltagirone, C., & the Early Diagnosis Group of the Italian Interdisciplinary Network on Alzheimer's Disease. (2005). Characterization of memory profile in subjects with amnestic mild cognitive impairment. *Journal of Clinical and Experimental Neuropsychology, 27(8),* 1033-1055.

Poon, L. W., & Fozard, J. L. (1980). Age and word frequency effects in continuous recognition memory. *Journal of Gerontology, 35*, 77-86.

Postman, L. (1970). Effects of word frequency on acquisition and retention under conditions of free-recall learning. *The Quarterly Journal of Experimental Psychology, 22*, 185-195.

Prince, M., Wimo, A., Guerchet, M., Ali, G. C., Wu, Y. T., & Prina, M. (2015). The Global Impact of Dementia: An analysis of prevalence, incidence, cost and trends. World Alzheimer Report. Retrieved online from https://www.alz.co.uk/research/WorldAlzheimerReport2015.pdf

Ralph, M. A. L., Graham, K. S., Ellis, A. W., & Hodges, J. R. (1998). Naming in semantic dementia—what matters?. *Neuropsychologia, 36(8),* 775-784.

Rasmusson, D. X., Bylsma, F. W., & Brandt, J. (1995). Stability of performance on the Hopkins Verbal Learning Test. *Archives of Clinical Neuropsychology, 10(1),* 21-26.

Rey, A. (1964). *L'examen clinique en psychologie*. Paris: Presses Universitaire de France.

Rieu, D., Bachoud-Lévi, A. C., Laurent, A., Jurion, E., & Dallabarba, G. (2006). Adaptation française du «Hopkins verbal learning test». *Revue neurologique, 162(6),* 721-728.

Roberts, W. A. (1972). Free recall of word lists varying in length and rate of presentation: A test of total-time hypotheses. *Journal of Experimental Psychology, 92*, 365-372.

Roodenrys, S., Hulme, C., Lethbridge, A., Hinton, M., & Nimmo, L. M. (2002). Word-frequency and phonological-neighborhood effects on verbal short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28,* 1019-1034.

Rubin, D. C. (1980). 51 properties of 125 words: A unit analysis of verbal behavior. *Journal of Verbal Learning and Verbal Behavior, 19(6),* 736-755.

Schrijnemaekers, A. M. C., de Jager, C. A., Hogervorst, E., & Budge, M. M. (2006). Cases with mild cognitive impairment and Alzheimer's disease fail to benefit from repeated

Grenfell-Essam, R., Hogervorst, E. & Rahardjo, T. B. W. (in prep)

exposure to episodic memory tests as compared with controls. *Journal of Clinical and Experimental Neuropsychology, 28(3),* 438-455.

Shapiro, A. M., Benedict, R. H. B., Schretlen, D., & Brandt, J. (1999). Construct and concurrent validity of the Hopkins Verbal Learning Test–revised. *The Clinical Neuropsychologist, 13(3),* 348-358.

Shepard, R. N. (1967). Recognition memory for words, sentences, and pictures. *Journal of Verbal Learning and Verbal Behavior, 6,* 156-163.

Shi, J., Tian, J., Wei, M., Miao, Y., & Wang, Y. (2012). The utility of the Hopkins Verbal Learning Test (Chinese version) for screening dementia and mild cognitive impairment in a Chinese population. *BMC neurology, 12:*136.

Silveri, M. C., Cappa, A., Mariotti, P., & Puopolo, M. (2002). Naming in patients with Alzheimer's disease: Influence of age of acquisition and categorical effects. *Journal of Clinical and Experimental Neuropsychology, 24(6)*, 755-764.

Sumby, W. H. (1963). Word frequency and serial position effects. *Journal of Verbal Learning and Verbal Behavior, 1*, 443-450.

Tierney, M. C., Nores, A., Snow, W. G., Fisher, R. H., Zorzitto, M. L., & Reid, D. W. (1994). Use of the Rey Auditory Verbal Learning Test in differentiating normal aging from Alzheimer's and Parkinson's dementia. *Psychological Assessment, 6(2),* 129-134.

Ward, G., Tan, L., & Grenfell-Essam, R. (2010). Examining the relationship between free recall and immediate serial recall: the effects of list length and output order. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*, 1207-1241.

Weingartner, H., Grafman, J., Boutelle, W., Kaye, W., & Martin, P. R. (1983). Forms of memory failure. *Science, 221,* 380-383.

Weingartner, H., Kaye, W., Smallberg, S. A., Ebert, M. H., Gillin, J. C., & Sitaram, N.
(1981). Memory failures in progressive idiopathic dementia. *Journal of Abnormal
Psychology, 90(3),* 187-196.

Welsh, K., Butters, N., Hughes, J., Mohs, R., & Heyman, A. (1991). Detection of abnormal
memory decline in mild cases of Alzheimer's disease using CERAD
neuropsychological measures. *Archives of Neurology, 48(3),* 278-281.

Wilson, R. S., Bacon, L. D., Fox, J. H., Kramer, R. L., & Kaszniak, A. W. (1983). Word
frequency effect and recognition memory in dementia of the Alzheimer type. *Journal
of Clinical and Experimental Neuropsychology, 5(2),* 97-104.

Wolk, D. A., Dunfee, K. L., Dickerson, B. C., Aizenstein, H. J., & DeKosky, S. T. (2011). A
medial temporal lobe division of labor: insights from memory in aging and early
Alzheimer disease. *Hippocampus, 21(5),* 461-466.

Woodard, J. L., Dunlosky, J., & Salthouse, T. A. (1999). Task decomposition analysis of
intertrial free recall performance on the Rey Auditory Verbal Learning Test in normal
aging and Alzheimer's disease. *Journal of Clinical and Experimental
Neuropsychology, 21(5),* 666-676.

Woods, S. P., Scott, J. C., Conover, E., Marcotte, T. D., Heaton, R. K., Grant, I., & HIV
Neurobehavioral Research Center (HNRC) Group. (2005). Test-retest reliability of
component process variables within the Hopkins Verbal Learning Test-Revised.
Assessment, 12(1), 96-100.

Xu, X, Rahardjo, T. B., Xiao, S., & Hogervorst, E. (2014). The Hopkins Verbal Learning
Test and Detection of MCI and Mild Dementia: A Literature Review. *Journal of
Alzheimer Disease and Parkinsonism, 4*:166.

Yesufu, A., Bandelow, S, Rahardjo, T. B., & Hogervorst, E. (2009). Chapter 8: Dementia
diagnoses. In the *Handbook of Cognitive Aging: Causes, Processes and Mechanisms*.

Grenfell-Essam, R., Hogervorst, E. & Rahardjo, T. B. W. (in prep)

Eds Quentin Gariépy and Raphaël Ménard . Nova Sciences Publishers: NY. Invited submission.

Acknowledgments

Funding

Appendix Captions

Appendix 1: Full demographic information for the age, gender and educational level matched participants in the Dementia and Control groups. Note: participants on the same row were matched.

Appendix 2: Pairwise comparisons, with a Bonferroni correction, to follow up the significant main effect of serial position for trial 1 (Figure 5, Panel A). The values show the significance level; significant values are indicated in bold.

Appendix 3: Pairwise comparisons, with a Bonferroni correction, to follow up the significant main effect of serial position for trial 2 (Figure 5, Panel B). The values show the significance level; significant values are indicated in bold.

Appendix 4: Pairwise comparisons, with a Bonferroni correction, to follow up the significant main effect of serial position for trial 3 (Figure 5, Panel C). The values show the significance level; significant values are indicated in bold.

Appendix 5: Pairwise comparisons, with a Bonferroni correction, to follow up the significant main effect of serial position for trials 1-3 (Figure 5, Panel D). The values show the significance level; significant values are indicated in bold.

Appendix 1.

Gender	Education level	Age (in years)	
		Control cases	Dementia cases
Female	Primary school	81	85
Female	Primary school	83	88
Female	Elementary school (unfinished)	60	60
Female	Elementary school (unfinished)	70	70
Female	Elementary school (unfinished)	77	77
Female	Elementary school (unfinished)	78	83
Female	Elementary school (unfinished)	66	66
Female	Elementary school (unfinished)	68	68
Female	Elementary school (unfinished)	81	85
Female	Elementary school	62	62
Female	Elementary school	74	74
Female	Elementary school	75	75
Female	Elementary school	75	76
Female	Elementary school	75	75
Female	Elementary school	85	85
Female	Elementary school	86	86
Female	High school	74	75
Female	No formal education	60	60
Female	No formal education	67	67
Female	No formal education	68	68
Female	No formal education	70	70
Female	No formal education	70	70
Female	No formal education	85	85
Female	No formal education	65	66
Female	No formal education	80	80
Female	No formal education	65	65
Female	No formal education	70	70
Female	No formal education	70	70
Female	No formal education	70	70
Female	No formal education	71	73
Female	No formal education	70	70
Female	No formal education	70	70
Female	No formal education	76	76
Female	No formal education	75	75
Female	No formal education	75	75
Female	No formal education	77	77
Female	No formal education	78	78
Female	No formal education	80	80
Female	No formal education	82	83
Male	Elementary school (unfinished)	60	60
Male	Elementary school (unfinished)	65	66
Male	Elementary school (unfinished)	68	69
Male	Elementary school (unfinished)	71	71
Male	Elementary school (unfinished)	76	79

Male	Elementary school (unfinished)	76	80
Male	Elementary school (unfinished)	77	80
Male	Elementary school (unfinished)	80	83
Male	Elementary school (unfinished)	82	87
Male	Elementary school	70	70
Male	Elementary school	72	72
Male	Elementary school	76	76
Male	Elementary school	76	78
Male	Elementary school	76	80
Male	Elementary school	80	85
Male	Elementary school	81	86
Male	No formal education	75	75
Male	No formal education	76	80
Male	No formal education	60	61
Male	No formal education	70	70
Male	No formal education	76	74
Male	No formal education	73	73
Male	No formal education	76	80

Appendix 2.

SP	2	3	4	5	6	7	8	9	10	11	12
1	**0.004**	1.000	**< 0.001**	**< 0.001**	**< 0.001**	**< 0.001**	**0.015**	1.000	**0.006**	1.000	0.241
2		0.133	1.000	1.000	1.000	0.065	1.000	1.000	1.000	**0.021**	1.000
3			**< 0.001**	**< 0.001**	**< 0.001**	**< 0.001**	0.459	1.000	0.099	1.000	1.000
4				1.000	1.000	1.000	1.000	**0.004**	1.000	**< 0.001**	0.215
5					1.000	1.000	0.573	**0.003**	1.000	**< 0.001**	0.163
6						1.000	0.671	**0.014**	1.000	**< 0.001**	0.322
7							**0.007**	**< 0.001**	**0.036**	**< 0.001**	**0.002**
8								1.000	1.000	0.091	1.000
9									1.000	1.000	1.000
10										**0.010**	1.000
11											0.189

Appendix 3.

SP	2	3	4	5	6	7	8	9	10	11	12
1	0.338	1.000	**< 0.001**	**< 0.001**	**< 0.001**	**< 0.001**	0.238	0.747	1.000	1.000	1.000
2		0.164	0.153	0.227	1.000	**< 0.001**	1.000	1.000	1.000	1.000	1.000
3			**< 0.001**	**< 0.001**	**< 0.001**	**< 0.001**	0.074	0.103	1.000	1.000	1.000
4				1.000	1.000	1.000	0.072	0.054	**0.009**	**0.001**	**0.002**
5					1.000	1.000	0.796	0.360	0.082	**0.012**	**0.019**
6						1.000	1.000	1.000	0.259	0.109	0.325
7							**0.006**	**0.001**	**< 0.001**	**< 0.001**	**< 0.001**
8								1.000	1.000	1.000	1.000
9									1.000	1.000	1.000
10										1.000	1.000
11											1.000

Appendix 4.

SP	2	3	4	5	6	7	8	9	10	11	12
1	1.000	1.000	**< 0.001**	**0.015**	0.702	**< 0.001**	0.723	1.000	1.000	1.000	1.000
2		1.000	**0.001**	0.086	0.245	**< 0.001**	0.859	0.881	1.000	1.000	1.000
3			**< 0.001**	**0.001**	**0.032**	**< 0.001**	0.078	**0.043**	1.000	1.000	1.000
4				1.000	1.000	1.000	1.000	1.000	0.146	**0.007**	**0.005**
5					1.000	1.000	1.000	1.000	1.000	0.241	0.117
6						0.059	1.000	1.000	1.000	1.000	1.000
7							0.054	**0.022**	**0.002**	**< 0.001**	**< 0.001**
8								1.000	1.000	1.000	1.000
9									1.000	1.000	1.000
10										1.000	1.000
11											1.000

Appendix 5.

SP	2	3	4	5	6	7	8	9	10	11	12
1	**0.007**	1.000	**< 0.001**	**< 0.001**	**< 0.001**	**< 0.001**	**0.001**	**0.005**	**0.044**	1.000	0.103
2		**0.030**	**0.010**	0.082	0.407	**< 0.001**	1.000	1.000	1.000	1.000	1.000
3			**< 0.001**	**< 0.001**	**< 0.001**	**< 0.001**	**0.003**	**< 0.001**	0.125	1.000	0.286
4				1.000	1.000	1.000	0.069	0.121	0.012	**< 0.001**	**0.005**
5					1.000	1.000	0.230	0.389	0.239	**< 0.001**	**0.048**
6						0.146	1.000	1.000	0.715	**0.006**	0.106
7							**< 0.001**	**< 0.001**	**< 0.001**	**< 0.001**	**< 0.001**
8								1.000	1.000	0.842	1.000
9									1.000	1.000	1.000
10										1.000	1.000
11											1.000

Table Captions

Table 1. Characteristics of Form 1 of the Hopkins Verbal Learning Test: word frequency, word length (number of syllables), age-of-acquisition, and orthographic neighbourhood size for both English and Indonesian languages.

Table 2. Summary of the ANOVA analyses conducted on the overall proportion of words recalled and the proportion of words recalled by category for trial 1, trial 2, trial 3 and overall for trials 1-3.

Table 3. The likelihood of accessing each category across trials 1-3 for the Control and Dementia cases.

Table 4. Summary of the ANOVA analyses conducted on the proportion of words recalled as a function of word frequency and orthographic neighbourhood size for trials 1-3.

Table 5. Summary of the ANOVA analyses conducted on the serial position curves for trial 1, trial 2, trial 3, and overall for trials 1-3.

Table 6. The learning and retention percentages across trials 2 and 3 for the Control and Dementia cases.

Table 7. Summary of the ANOVA analyses conducted on the proportion of re-remembered, new, forgotten, and never recalled words.

Table 1.

		English version				Indonesian version			
	Words	Word frequency (per million)	Word length (syllables)	Age of acquisition (years)	Orthographic neighbourhood size	Words	Word frequency	Word length (syllables)	Orthographic neighbourhood size
1	lion	15.35	1	4.42	4	singa	923	2	1
2	emerald	2.57	3	8.26	0	intan	21	2	3
3	horse	92.88	1	4.15	8	kuda	3584	2	4
4	tent	17.49	1	5.16	15	tenda	485	2	5
5	sapphire	1.20	2	9.22	0	akik*	1	2	4
6	hotel	103.22	2	6.05	2	hotel	3160	2	1
7	cave	13.98	1	6.74	17	gua	1513	2	4
8	opal	1.02	2	9.28	2	kecubung*	3	3	0
9	tiger	18.53	2	4.00	3	harimau	583	3	0
10	pearl	15.67	1	6.28	1	mutiara	305	4	0
11	cow	25.51	1	3.94	24	sapi	1401	2	11
12	hut	13.22	1	8.10	17	gubuk	144	2	4
Average (SD)	Overall	26.72 (32.75)	1.50 (0.65)	6.30 (1.94)	7.75 (7.94)		1010.25 (1168.21)	2.33 (0.62)	3.08 (2.98)
	Animals	38.07 (31.86)	1.25 (0.43)	4.13 (0.19)	9.75 (8.44)		1622.75 (1169.02)	2.25 (0.84)	3.20 (4.30)
	Stones	5.12 (6.12)	2.00 (0.71)	8.26 (1.21)	0.75 (0.83)		82.50 (128.70)	2.75 (0.83)	1.40 (1.74)
	Shelter	36.98 (38.28)	1.25 (0.43)	6.51 (1.07)	12.75 (6.26)		1325.50 (1172.92)	2.00 (0.00)	2.80 (1.50)

Note: *denotes non-direct translations: akik = agate, kecubung = amethyst.

Table 2.

	df	MSE	F	p
*Overall proportion of words recalled**				
Group	1, 122	10.2	69.9	<.001
Trial	1.785, 217.7	1.46	25.8	<.001
Group x trial	1.785, 217.7	1.46	11.2	<.001
*Proportion of words recalled by category for trial 1**				
Group	1, 122	0.073	37.5	<.001
Category	1.904, 232.3	0.046	49.7	<.001
Group x category	1.904, 232.3	0.046	6.01	0.003
Proportion of words recalled by category for trial 2				
Group	1, 122	0.082	61.0	<.001
Category	2, 244	0.044	31.0	<.001
Group x category	2, 244	0.044	1.43	0.243
*Proportion of words recalled by category for trial 3**				
Group	1, 122	0.110	69.3	<.001
Category	1.897, 231.4	0.044	17.9	<.001
Group x category	1.897, 231.4	0.044	0.593	0.554
Proportion of words recalled by category for trials 1-3				
Group	1, 122	0.154	66.0	<.001
Category	2, 244	0.049	43.1	<.001
Group x category	2, 244	0.049	0.807	0.447

Note: * indicates the initial assumption checks indicated that Mauchly's Test of Sphericity was violated therefore degrees of freedom were corrected using the Greenhouse-Geisser correction.

Table 3.

	Chi-square			t-test	
	χ^2	p	df	t	p
Trial 1					
Animal	7.49	0.006	93	2.96	0.004
Stones	7.64	0.006	59	1.32	0.191
Shelter	8.18	0.004	55	0.34	0.739
Trial 2					
Animal	6.86	0.009	92	2.78	0.007
Stones	9.76	0.002	76.7	4.02*	< 0.001
Shelter	11.48	<0.001	54.7	2.79*	0.007
Trial 3					
Animal	8.49	0.004	90	3.30	0.001
Stones	12.68	<0.001	84	3.37	0.001
Shelter	11.69	<0.001	71	3.45	0.001

Note: * indicates the initial assumption checks indicated that Levene's Test for Equality of Variances was violated therefore the equal variance not assumed values were reported.

Table 4.

	df	MSE	F	p
Word frequency				
Group	1, 122	0. 141	69.91	<.001
Word frequency	1, 122	0.054	11.77	.001
Trial*	1.785, 217.7	0.018	25.84	<.001
Group x word frequency	1, 122	0.054	0.07	.792
Group x trial	2, 244	0.018	11.22	<.001
Word frequency x trial	2, 244	0.017	5.59	.004
Group x word frequency x trial	2, 244	0.017	2.20	.114
Orthographic neighbourhood size				
Group	1, 122	0. 141	69.91	<.001
Orthographic N	1, 122	0.046	5.97	.016
Trial*	1.785, 217.7	0.018	25.84	<.001
Group x orthographic N	1, 122	0.046	2.72	.102
Group x trial	2, 244	0.018	11.22	<.001
Orthographic N x trial	2, 244	0.023	0.36	.701
Group x orthographic N x trial	2, 244	0.023	0.37	.694

Note: * indicates the initial assumption checks indicated that Mauchly's Test of Sphericity was violated therefore degrees of freedom were corrected using the Greenhouse-Geisser correction.

Table 5.

	df	MSE	F	p
Serial position curve for trial 1*				
Group	1, 122	0. 293	37.5	<.001
Serial position	9.214, 1124.2	0.187	12.8	<.001
Group x serial position	9.214, 1124.2	0.187	1.90	0.036
Serial position curve for trial 2*				
Group	1, 122	0.330	61.0	<.001
Serial position	9.694, 1182.7	0.194	10.8	<.001
Group x serial position	9.694, 1182.7	0.194	1.30	0.217
Serial position curve for trial 3*				
Group	1, 122	0.440	69.3	<.001
Serial position	9.493, 1158.2	0.200	8.53	<.001
Group x serial position	9.493, 1158.2	0.200	1.25	0.248
Serial position curve for trials 1-3*				
Group	1, 122	0.616	66.0	<.001
Serial position	9.581, 1168.9	0.191	16.3	<.001
Group x serial position	9.581, 1168.9	0.191	0.929	0.511

Note: * indicates the initial assumption checks indicated that Mauchly's Test of Sphericity was violated therefore degrees of freedom were corrected using the Greenhouse-Geisser correction.

Table 6.

	SP	1	2	3	4	5	6	7	8	9	10	11	12	Total
	Word	*Lion*	*Emerald*	*Horse*	*Tent*	*Sapphire*	*Hotel*	*Cave*	*Opal*	*Tiger*	*Pearl*	*Cow*	*Hut*	
						CONTROL GROUP								
Trial 2	Re-remembered	0.37	0.26	0.44	0.11	0.13	0.13	0.06	0.19	0.39	0.24	0.35	0.24	0.24
	New	0.21	0.23	0.21	0.13	0.10	0.13	0.11	0.26	0.10	0.23	0.11	0.26	0.17
	Forgotten	0.16	0.08	0.13	0.08	0.05	0.05	0.03	0.13	0.13	0.03	0.19	0.10	0.10
	Never recalled	0.26	0.44	0.23	0.68	0.73	0.69	0.79	0.42	0.39	0.50	0.34	0.40	0.49
Trial 3	Re-remembered	0.50	0.48	0.53	0.18	0.18	0.21	0.10	0.35	0.40	0.35	0.48	0.53	0.36
	New	0.11	0.15	0.10	0.10	0.19	0.24	0.11	0.10	0.08	0.15	0.08	0.10	0.13
	Forgotten	0.24	0.08	0.24	0.15	0.10	0.10	0.11	0.23	0.21	0.15	0.18	0.06	0.15
	Never recalled	0.15	0.29	0.13	0.58	0.53	0.45	0.68	0.32	0.31	0.35	0.26	0.31	0.36
						DEMENTIA GROUP								
Trial 2	Re-remembered	0.23	0.06	0.13	0.02	0.06	0.08	0.02	0.06	0.10	0.06	0.19	0.15	0.10
	New	0.13	0.08	0.19	0.03	0.05	0.06	0.02	0.08	0.05	0.16	0.10	0.08	0.09
	Forgotten	0.13	0.05	0.11	0.05	0.02	0.02	0.03	0.11	0.05	0.10	0.13	0.06	0.07
	Never recalled	0.52	0.81	0.56	0.90	0.87	0.84	0.94	0.74	0.81	0.68	0.58	0.71	0.75
Trial 3	Re-remembered	0.23	0.13	0.32	0.05	0.06	0.11	0.03	0.16	0.08	0.18	0.24	0.15	0.15
	New	0.05	0.11	0.05	0.08	0.06	0.03	0.03	0.00	0.06	0.05	0.03	0.06	0.05
	Forgotten	0.26	0.06	0.11	0.05	0.06	0.05	0.03	0.10	0.11	0.15	0.18	0.15	0.11
	Never recalled	0.47	0.69	0.52	0.82	0.81	0.81	0.90	0.74	0.74	0.63	0.55	0.65	0.69

Guide to categories: *Re-remembered* means that it was recalled on the current trial and had been recalled on previous trial(s), *New* means that it was recalled on the current trial but had not been recalled on any previous trial(s), *Forgotten* means that it was not recalled on the current trial but it had been recalled on previous trial(s), *Never* recalled means that it was not recalled on the current trial and had not been recalled on any previous trial(s).

Table 7.

	df	MSE	F	p
Proportion of re-remembered words				
Group	1, 122	0.421	57.3	<.001
Trial	1, 122	0.101	49.5	<.001
Group x trial	1, 122	0.101	8.32	0.005
Proportion of new words				
Group	1, 122	1.51	37.3	<.001
Trial	1, 122	1.20	12.1	0.001
Group x trial	1, 122	1.20	0.337	0.563
Proportion of forgotten words				
Group	1, 122	2.49	4.39	0.038
Trial	1, 122	0.709	27.9	<.001
Group x trial	1, 122	0.709	1.12	0.293
Proportion of never recalled words				
Group	1, 122	12.7	60.7	<.001
Trial	1, 122	0.500	14.1	<.001
Group x trial	1, 122	0.500	23.5	<.001

Figure Captions

Figure 1. Mean proportion of words recalled overall for trials 1-3 in the HVLT for control and dementia cases.

Figure 2. Mean proportion of words recalled within each category of the HVLT for control and dementia cases for trial 1 (A), trial 2 (B), and trial 3 (C). Panel D shows the overall likelihood of recalling the words within the category across the three trials.

Figure 3. Percentage of participants accessing each category for trial 1 (A) trial 2 (B), and trial 3 (C) and average number of words recalled given the category was accessed for trial 1 (D), trial 2 (E), and trial 3 (F) for control and dementia cases.

Figure 4. Mean proportion of words recalled for trials 1-3 in the HVLT for control and dementia cases for high and low frequency words (A) and large and small orthographic neighbourhood size words (B).

Figure 5. Overall serial position curves for control and dementia cases for trial 1 (A), trial 2 (B), and trial 3 (C). Panel D shows the overall likelihood of recalling each word across the three trials.
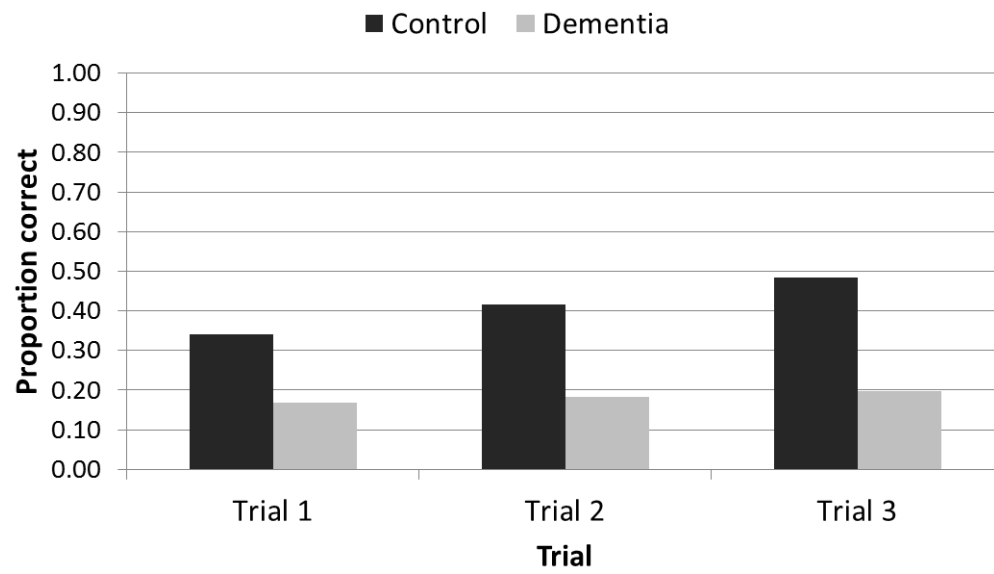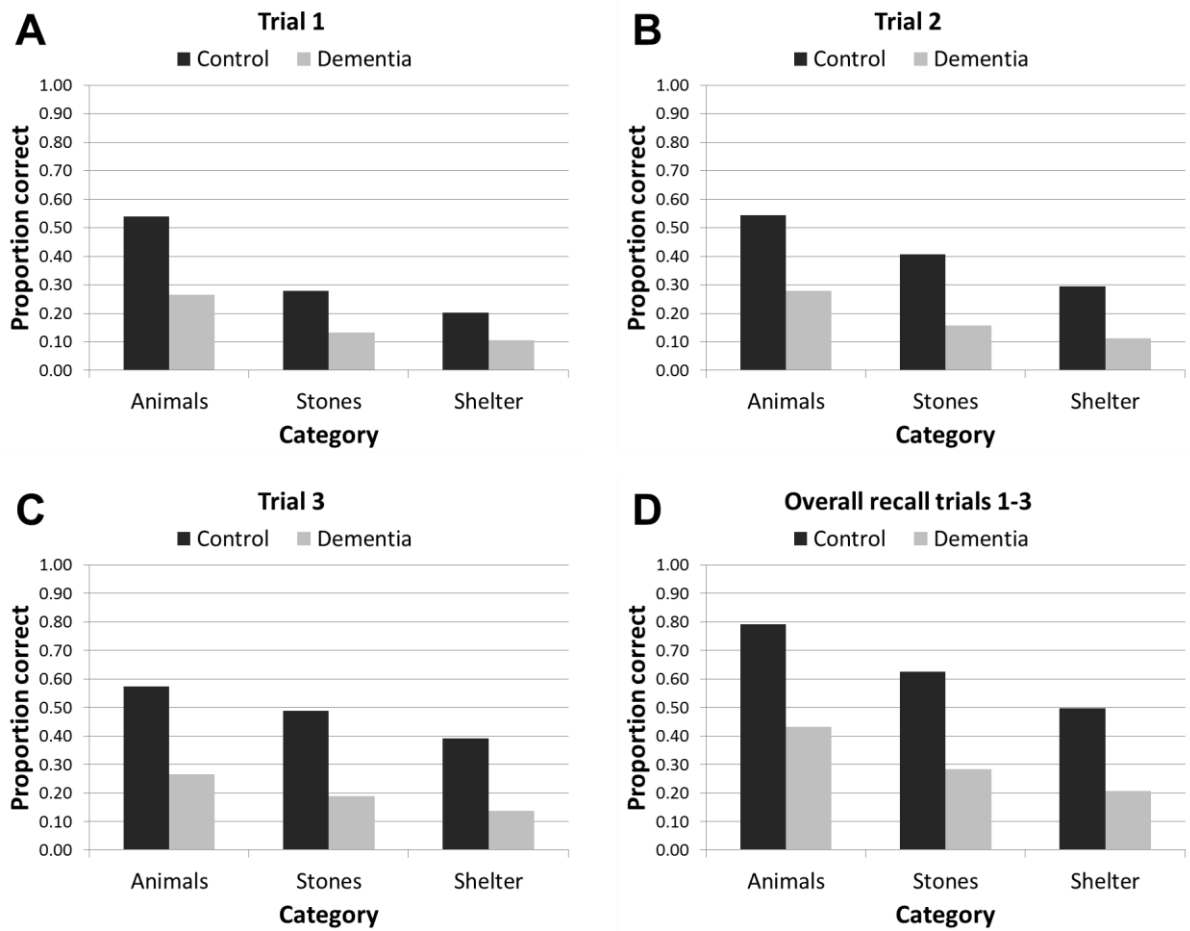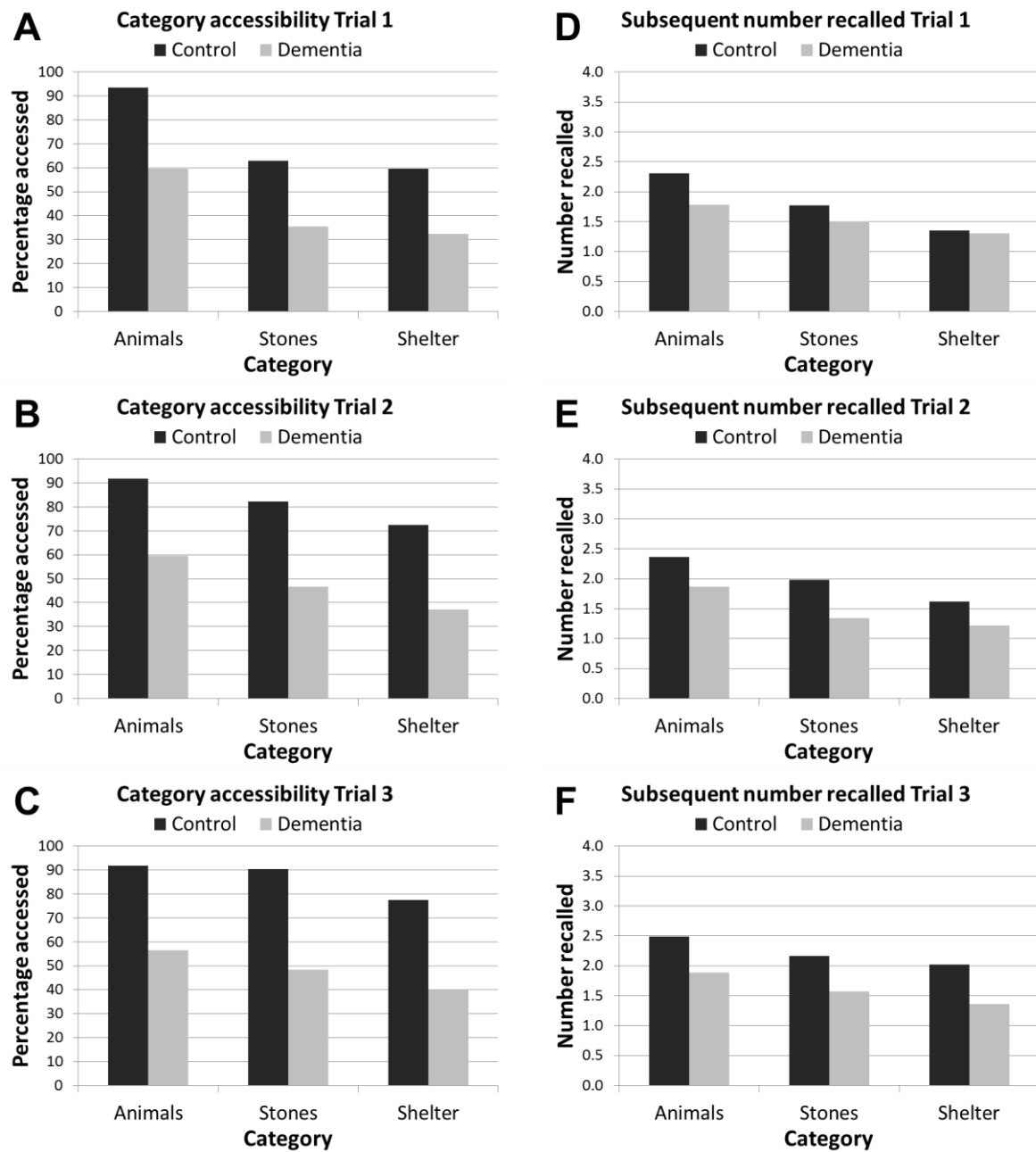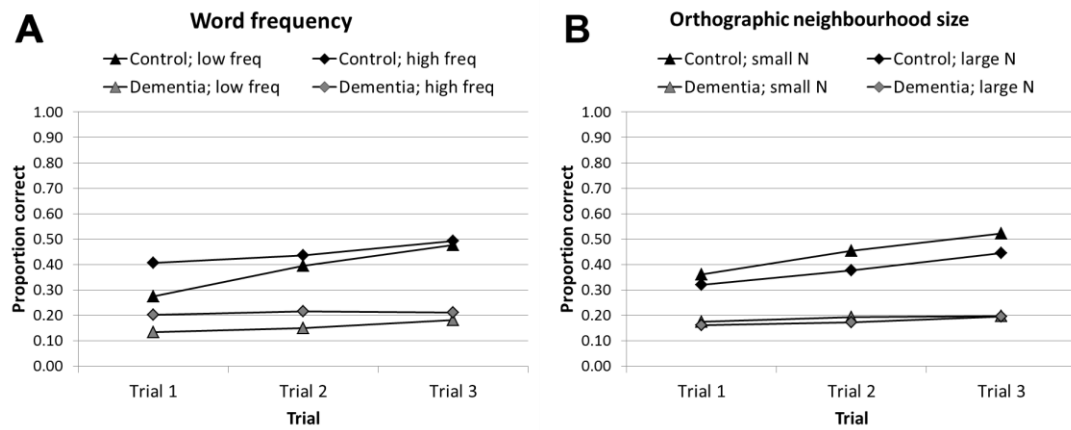
Figure 1.

Figure 2.

Figure 3.

Figure 4.

Figure 5.