Abstract

Objective: Most existing tests of memory and verbal learning in adults were created for spoken languages, and are unsuitable for assessing deaf people who rely on signed languages. In response to this need for sign language measures, the British Sign Language Verbal Learning and Memory Test (BSL-VLMT) was developed. It follows the format of the English language Hopkins Verbal Learning Test Revised (Benedict, Schretlen, Gronninger & Brandt, 1998) using standardised video presentation with novel stimuli and instructions wholly in British Sign Language, and no English language requirement.

Method: Data were collected from 223 cognitively-healthy deaf signers aged 50-89 and 12 deaf patients diagnosed with dementia. Normative data percentiles were derived for clinical use, and Receiver Operating Characteristic (ROC) curves computed to explore the clinical potential and diagnostic sensitivity and specificity.

Results: The test showed good discrimination between the normative and clinical samples, providing preliminary evidence of clinical utility for identifying learning and memory impairment in older deaf signers with neurodegeneration.

Conclusions: This innovative video testing approach transforms the ability to accurately detect memory impairments in deaf people and avoids the problems of using interpreters, with international potential for adapting similar tests into other signed languages.

Introduction

Verbal learning memory tasks such as the California Verbal Learning Test (CVLT, Delis, Kramer, Kaplan & Ober, 1987) and the Hopkins Verbal Learning Test- Revised (HVLT, Brandt, 1991, HVLT-R, Benedict et al., 1998) are often used to identify memory impairments and neurodegeneration in users of spoken languages. These tests are unsuitable for deaf people who use signed languages. Verbal learning tests derived directly in sign language to date are limited to American Sign Language (ASL) and include the Signed Paired Associates Test (Pollard, Rediess & DeMatteo, 2005) which is analogous to the Verbal Paired Associates subtest of the Wechsler Memory Scale-Revised (WMS-R), 1987; and the Signed Verbal Learning Test, (Morere, 2013) which is loosely based on the CVLT. The stimuli and instructions for these measures are published in written English, with no standardised, video-presentation available in ASL. We present the first test of verbal learning and memory in British Sign Language (BSL) based on the format of the HVLT-R. The BSL-Verbal Learning and Memory Test (BSL-VLMT) was developed as a verbal memory measure with wholly signed instructions and stimuli in video format with no spoken language requirement. We adopted the format of the HVLT-R with permission from the authors; however, our test is not a translation and uses completely novel signed stimuli. This paper describes the development and validation of this new test.

Signed languages like BSL, which is used in the United Kingdom, are independent languages, unrelated to spoken languages with their own lexicons, grammars (Sutton-Spence & Woll, 1999) and unique properties that must be considered when developing or translating tests. Lexical signs can be described in terms of their phonological structure in terms of their handshape, movement and location

(MacSweeney et al., 2006), and often show greater iconicity than spoken words (i.e. the form of the sign may be influenced by the form of the object to which it refers). Some classes of signs share similar form and action with gestures, particularly those relating to tool use, body parts or human action (Sutton-Spence & Woll, 1999). Care must be taken when selecting test stimuli to take into account lexical variables such as phonological similarity, gesture similarity and iconicity. Within BSL, dialect differences in the lexicon related to geographical region and the signer's age are common (Stamp et al., 2011), so tests must use familiar and widely understood vocabulary.

The intrinsic differences between spoken and signed languages makes accurate translation of existing memory tests difficult and scores unreliable, increasing the risk of clinical misjudgement (Vernon & Miller, 2001; Cornes & Napier, 2005). Tests often have linguistic and cultural components, assumed knowledge and semantic relationships based on spoken languages that do not translate well. Using sign language interpreters may result in a loss of control over the psychological properties of the stimuli. For example, verbal memory may be impeded by unconsidered phonological similarity between translations of the stimuli (Wilson & Emmorey, 1997), or the original semantic groupings of test items may not be relevant once translated because of language differences in semantic associations. For example, a signer may group HAT with other things to do with the head, rather than other items of clothing. Validity is further compromised and scores made clinically meaningless by the lack of deaf normative data. Additionally, poor literacy levels among deaf people (Powers, Gregory & Thoutenhoofd, 1999) prevent modified administration using written versions of spoken language instructions because this would measure acquired reading ability rather than innate cognitive ability. For these reasons, cognitive assessment of deaf signers is often limited to the nonverbal domain (Baker & Baker, 2011) because there is broad clinical consensus that, even when translated, spoken language tests measuring verbal cognition, such as the Wechsler Memory Scale-Fourth Edition (WMS-IV; Wechsler, 2009) or verbal subtests of the Wechsler Adult Intelligence Scale–Fourth Edition (WAIS-IV; Wechsler, 2008) are unsuitable (Pollard, 2002). This means it is difficult to confidently identify conditions where diagnosis is informed by specific patterns of verbal memory impairment, or the contrast between verbal and nonverbal cognition (Pollard, Rediess & DeMatteo, 2005).

This paper describes a new verbal memory test for users of BSL. The HVLT-R format was chosen, as it is user friendly, simple to administer via video, and has been widely adapted into spoken languages other than English (French, Rieu, Bachoud-Levi, Laurent, Jurion & Dalla Barba, 2006, Spanish, Cherner, Suarez, Lazzaretto, Fortuny, Rivera, Mindt et al., 2007 and Chinese, Shi, Tian, Wei, Miao &Wang, 2012). It has a solid documented clinical utility for a range of neurological disorders (Frank & Byrne, 2000). It provides high diagnostic sensitivity and specificity for Alzheimer's disease (Brandt, 1991; Shapiro, Benedict, Schretlen & Brandt, 1999); it has also been used with many other patient groups including patients with vascular dementia (Frank & Byrne, 2000; Hogervorst, Combrinck, Lapuerta, Rue, Swales & Budge, 2002), mild traumatic brain injury (Bruce & Echemendia, 2003) and HIV associated neurocognitive disorders (Woods, Cobb Scott, Dawson, Morgan, Carey, Heaton et al., 2005). The HVLT-R shows convergent validity and robust correlation with other measures of verbal memory including the CVLT (Delis, Kramer, Kaplan & Ober, 1987) and WMS-R Logical Memory subtest (Wechsler, 2009). There is also proven reliability between alternative forms of the test that use different stimuli (Benedict et al., 1998), suggesting that the format is likely to be adaptable to using stimuli from different languages, including signed languages.

The original English language HVLT-R test is unsuitable for deaf signers as the psychological integrity of the test is compromised by translation. Exemplars from semantic categories such as precious stones or minerals used in the HVLT-R are typically fingerspelled in BSL or conveyed via a generic sign for jewellery stones (i.e. a sign indicating a small round object on the ring finger) accompanied by mouthing of the equivalent English word (‘emerald’, ‘sapphire’, ‘opal’ and ‘pearl’). These category exemplars make poor candidates for BSL lexical recall because they do not have distinct lexical signs and borrow heavily from English. For such reasons, we developed completely novel stimuli with BSL as our starting point.

The new BSL-VLMT replicates the format of the HVLT-R. The test comprises a 12-item list of BSL signs, drawn from three semantic categories, that respondents must immediately recall, in any order, after each of three learning trials. This is followed by delayed free recall after a 20-25 minute interval during which other tests are undertaken, and a final yes/no recognition trial made up of the 12 target signs and 12 distractor signs, half of which are semantically related and half unrelated.

The BSL-VLMT like the HVLT-R produces the following key measures:

a) **Immediate Recall**

   i) Immediate Recall (sum of trials 1-3)

   ii) Learning Index –a measure of learning across trials 1-3 (higher of trial 2 or 3 minus trial 1)

b) **Delayed Recall**

   i) Delayed Recall (trial 4)

   ii) Retention Index which measures the percentage of items retained from earlier learning trials (trial 4 divided by higher of trials 2 or 3 multiplied by 100)

c) **Recognition**

   i) Recognition score (No. of correct true positives)

   ii) False Positive Errors

   iii) Discrimination Index (true positives minus false positives).

The score profile provides information about which type of memory is problematic. Poor Immediate Recall and Learning Index scores suggest problems with encoding new information, which impacts the formation of new memories and learning. Delayed Recall provides a measure of retrieval memory or the ability to access previously stored information. Where a person is able to form memories but has specific difficulties with memory retrieval we would expect Recognition to be normal, and conversely Recognition would be impaired where there was deficient encoding.

Here we describe test development and participant variables that affect performance. We report test norms, validation and clinical sensitivity in distinguishing twelve deaf people with dementia from healthy controls.

Methods

Test design

The BSL-VLMT uses the HVLT-R format with novel BSL stimuli selected from Vinson, Cormier, Denmark, Schembri & Vigliocco's (2008) lexical norming study. Stimuli were derived from three carefully matched semantic categories (food, animals and clothing) with no statistically significant differences in iconicity, familiarity or age of acquisition between the categories. An additional 12 foil items for the recognition trial were drawn from names of countries, sports and transport. There were no differences in lexical variables between target and foil groups, with the exception of significantly higher age of acquisition for countries than animals F(5, 24)=3.7, MSE=10.5, p=.016 η^2= .5). Stimuli were assigned to target or foil lists following the procedure reported for HVLT-R (Brandt, 1991), with the two most familiar exemplars used as foils and the next four serving as targets. All stimuli have low regional variation and are well understood across the UK. Synonyms were excluded to ensure that signers recalled the target lexical item and not a sign with a related meaning or a visually similar sign with a different meaning (Vinson et al., 2008).

BSL-VLMT instructions and stimuli were presented in BSL video format viewed on a laptop. Stimuli were presented by an older deaf native signer at a rate of one every 2 seconds. The use of an older signer ensured ecological validity by using signs familiar to older deaf people. Between each sign the model returned her hands to a resting position on her lap, similar to pausing between items when reading a list of words aloud.

Participants

Normative data

Normative data were collected from 223 (80 male, 143 female) cognitively healthy deaf participants aged 50-89 years (M 68.26, SD 10.24), attending holiday camps or social clubs for older deaf people in the South of England. All participants were users of British Sign Language and were born deaf or had lost their hearing before the age of 10 years, an age cut-off that includes the majority of people who use sign language as a first or preferred language. We wanted a broad sample to include those who had been deafened due to childhood illness. It was important that the normative data reflected the diversity of the UK Deaf population, which is heterogeneous in nature with differences in language development and proficiency, educational background, and age of deafness onset. Applying more restrictive exclusion criteria would run the risk of testing the measure on an narrow section of British sign language users, and hence reducing the clinical utility of the data.

A screening interview ensured inclusion criteria were met, with no known neurological history, visual impairment, additional disability or substance abuse. This interview also collected data on: age, the history and cause of deafness, age of sign language acquisition, education and occupation history. A team of 13 specially trained deaf investigators administered the BSL-VLMT as part of a broader battery of cognitive tests, which included the BSL Cognitive Screening Test (BSL-CST, norms are reported in Atkinson, Denmark, Marshall, Mummery & Woll, 2015) and the Modified Digit Span (using numerals) adapted by our team for use in a pilot cognitive disorders clinic for deaf patients, which will be reported elsewhere. Matrix Reasoning t score was used as a measure of nonverbal intellectual ability (M 52.11, SD 10.91, WAIS-IV; Wechsler, 2008). A composite language score was computed from comprehension, production and lexical naming items from the cognitive screen to enable consideration of the influence of verbal ability on BSL-VLMT performance.

Clinical data

Recruitment took place at the National Hospital for Neurology and Neurosurgery (NHNN), London, where we piloted a monthly cognitive disorders clinic for deaf patients using our new battery of tests as part of diagnostic workup, in the absence of other suitable tests for deaf signers. BSL-VLMT data were collected prior to clinical consultation and diagnosis. Consensus diagnosis by a multidisciplinary panel employed standard clinical diagnostic criteria for dementia (DSM-IV TR, American Psychiatric Association, 2000) and disease specific criteria (e.g. NINCDS-ADRDA McKhann et al., 1984; Dubois, Feldman, Jacova, Dekosky, Barberger-Gateau et al., 2007 for AD; Neary, Snowden, Gustafson, Passant, Stuss et al, 1998 for FTLD). Diagnosis was based on clinical assessment and history, supported by the results of imaging, neurophysiology, and immunological tests as well as the BSL-Cognitive Screening Test results.

Data for twelve patients (six female) with clinically identified dementia are reported. Nine individuals had a clinical diagnosis of Alzheimer’s disease, one frontotemporal dementia, one genetic non-AD dementia likely to be secondary to mitochondrial disease, and one amnestic mild cognitive impairment. Table 1 shows demographic information for normative and dementia groups.

Procedure

Following protocol similar to the HVLT-R, our participants were required to complete three learning trials for immediate recall, with delayed recall and recognition trials taking place after a 20-25 minute delay, during which other tests were completed (Matrix Reasoning and Modified Digit Span). Responses were recorded using an English gloss of the BSL sign or circling recognition items on a score sheet, and were additionally filmed for later verification due to occasional difficulties in writing responses while watching the participant signing. The first author calculated scores for the whole sample and a second rater re-scored 20% of the sample (n=45) to check inter-rater reliability. For each participant the number of semantic clusters during recall was coded (i.e. the number of items recalled adjacent to others from the same category: food, animals and clothing).

Ethical approval was obtained from UCL Graduate School Ethics Committee and informed consent was obtained from all participants. Clinical patients additionally agreed that anonymised scores collected during routine care could be used for research purposes.

Analysis

Normative data were used to establish test validity and reliability as a measure of verbal memory and learning in deaf people. Inter-rater reliability was assessed using intraclass correlation to compare the BSL-VLMT scores obtained by the two independent scorers. Convergent validity is hard to establish because there are no other existing measures of verbal memory in BSL, so we report Pearson correlation coefficients for the verbal learning component of the BSL-CST where respondents have to recall biographical information about a deaf man (Atkinson et al, 2015). We also report correlations with a composite score of verbal ability. For the normative control sample, Pearson correlation coefficients were calculated for immediate recall scores and variables showing normal distribution including: age; nonverbal ability; age of BSL acquisition; and years of education. A Mann-Whitney U test was used for the non-normally distributed variable - age of deafness onset. Differences in immediate recall scores between demographic subgroups including: gender; native, early or late BSL acquisition; cause of deafness; occupation; educational attainment (the highest educational credential obtained): and academically selective schooling (based on grammar school attendance which was historically determined by entrance exams in the UK), were examined using independent sample t-tests and one-way analyses of variance. Multiple regression was used to examine which demographic variables explained variance in immediate recall performance.

Percentiles for normative performance were generated for clinical comparison. Mann Whitney-U tests were used to confirm demographic similarity between control and dementia groups, and to examine differences in performance on each BSL-VLMT measure. Areas under the receiver-operator curves (ROCs) were calculated to establish clinical sensitivity and specificity; and to determine how accurately each distinguishes patients with dementia from controls.

Reliability and validity

Inter-rater reliability An intraclass correlation of .99 (p<.0001) showed excellent inter-rater reliability with very little discrepancy between first and second raters.

Convergent validity Immediate recall score on the BSL-VLMT (sum of trials 1-3) showed a significant relationship with immediate recall on the only other BSL test of verbal learning and memory, the Deaf Man Verbal Learning task, which forms part of the BSL-Cognitive Screening Test. Correlation controlling for age showed a medium-sized effect of convergence between scores on the two tests r(221)= .352, p<0.001. Verbal ability on the BSL-CST also positively correlated with immediate recall performance r(221)=.362, p<0.001, with those with higher ability remembering more items, but verbal ability showed no significant relationship to the amount improvement across the three learning trials r(221)=.060, p=.369

Relationship to demographic variables

Table 2 shows correlations between BSL-VMLT immediate recall and demographic variables. Test scores did not significantly differ between groups based on UK region, age became deaf, age of BSL acquisition, cause of deafness or occupational status. There was a small effect of age with a significant decrease in mean performance among older participants. There was also a significant relationship showing that better test scores corresponded with higher nonverbal intellectual ability (measured by Matrix Reasoning). Linear multiple regression with simultaneous entry of age, gender, years of education, educational attainment and nonverbal ability, showed that only age and gender explain a significant amount of the variance in immediate recall scores F(5,218)=8.84, p<.001 R square =.17; Adjusted R square=.15. Although, there was a small to moderate effect of length of education and a negligible effect of educational attainment, this was fully accounted for by their association with age, which emerged as the strongest predictor (β = -.16, p <.001). Gender showed a small effect on immediate recall with women performing slightly better than men (β = 1.39, p <.05) once other factors had been taken into account.

Normative data for clinical use

Immediate recall

Age was the most robust factor influencing immediate recall, so percentiles for each age band within the whole sample are reported in Table 3 enabling clinicians to compare obtained scores to the normative range for the patient's age group. Table 4 provides norms for the learning index which are collapsed across age-bands because there was no detectable correlation between age and the amount of learning across immediate recall trials r(221)=0.001, p=.987.

Delayed recall

Delayed recall was a challenging task with just over a third (34.2%, n=79) of the normative sample unable to recall any items. However, despite difficulty with the delayed recall task there was no significant difference between this sub-group and the rest of the sample in the rate of learning across trials t(221)=1.034, p=.302 or recognition scores t(221)=1.563, p=.120. Older age groups were more affected (50-59: 19.6%; 60-69: 30.3%; 70-79: 42%; and 80-89: 51.0%). Most of those scoring zero for delayed recall had maximum scores for recognition (74.7%, n=59), made no false identification of foils (70.9%, n=56), and had immediate recall scores which indicated no difficulty with encoding. Those scoring zero: recalled fewer items across the learning trials t(221)=5.347, p<.001 (a mean of four items less across the three trials); had lower composite score on language items in the BSL Cognitive Screening Test t(221)=4.595, p<.001 and lower nonverbal intellectual ability as measured by Matrix Reasoning t(221)=4.205, p<.001. The finding of normal rates of learning and recognition memory in this group suggests that retrieval problems may be at least in part due to the task demands, which require good language comprehension of instructions.

The large number of zero scores mean that the distribution of normative scores delayed recall was highly skewed. For this reason, norms are reported in Table 5 for only for the two-thirds of the control sample who were able to recall at least one item (n=144). This smaller sample has a normal distribution which enabled percentiles to be extracted. Within this group 96.53% (n=139) scored 6 or more. The fewest items recalled was 5 for all age-groups except the oldest participants, aged 80-89 years.

Suggesting that where task instructions were understood retrieval was robust. However, delayed recall percentiles must be used with caution by clinicians because they are based on a higher ability subgroup, and a score of zero on delayed recall is not necessarily pathological.

Recognition

This task is designed as a screen and has a low ceiling. Taking the whole of the normative sample, the majority obtained the maximum score for recognising test items (79%) and only 6.3% made 2 or more errors. Recognition data is not normally distributed so percentile cut offs are provided for scores falling below ceiling in Table 6. False positive error cut offs and discrimination index percentiles are also reported in Table 6. These norms are collapsed across age-bands because there were no significant age differences.

Clinical data

The dementia group was significantly older than the normative sample, as all but two participants were aged 70-89 years. To enable comparison within the same age range we used Mann-Whitney U tests to compare the BSL-VLMT scores of the 10 people with dementia aged 70-89 years to a subgroup of 95 controls who were also aged 70-89 years. Results on Table 7 show that these groups were well matched demographically, with no significant differences in age, years of education, or age of BSL acquisition. Statistical comparison of BSL-VLMT scores showed patients with dementia achieved a consistently lower distribution of scores than controls on immediate recall and learning across trials 1-3 as measured by the learning index. On these trials, they made significantly greater numbers of intrusion errors and showed less semantic clustering in their responses. Clustering remained significant when a ratio score between number of clusters and total number of items produced was used t(232)=4.934, p<.001. The dementia group showed significantly poorer performance for delayed recall, and poorer retention index scores, indicating fewer items retained between immediate and delayed trials, and significantly more false positive errors on the recognition task, wrongly identifying foils as having been seen earlier. There was also a significant difference in the discrimination index, which measures the difference between correct recognition responses and false positive errors. These results lead to the rejection of the null hypothesis that distributions of BSL-VLMT scores are the same across healthy and patient groups. There are two exceptions, firstly, there was no statistical difference in the number of repetition errors during immediate recall and secondly, there were no group differences in recognition of items seen in earlier trials.

ROC curves were computed for all key measures showing significant group differences, to assist in interpretation of diagnostic utility. It should be noted that the small dementia sample size mean that these findings are exploratory and should only be used as a guide. The areas under the curves (AUC) for the whole dementia and normative samples aged 50-89 years are reported in Table 8 showing the ability of each measure to reliably distinguish between deaf patients with dementia and controls. The true positive rate was the percentage of dementia patients correctly classified as having dementia using BSL-VLMT score. This was plotted against the false positive rate, which shows the percentage of controls misclassified as belonging to the dementia group. Tradeoffs between sensitivity and specificity are also reported in the table using a rule of thumb of >.9 excellent, .8 -.9 good, .7 -.8 fair and .6 -.7

poor. Taking one example, the AUC for immediate recall was 0.945 (94.5%, P<.001, CI=95%) indicating that 94.5% of patients were correctly classified as belonging to the dementia group with a low false positive rate, with few controls erroneously classified as having dementia. This AUC value is high, indicating reliable clinical accuracy, whereas a value of .50 would indicate the predictor is no better than chance (Zhou & Obuchowski, Obushcowski, 2002). Immediate recall, number of false positive recognition errors and the discrimination index showed the greatest potential for diagnostic reliability with an excellent tradeoff between sensitivity and specificity. These three measures are discussed in more detail below.

Immediate recall

Our dementia sample is too small to establish an absolute cut-off score for immediate recall below which there would be a very high chance that a person has dementia or another memory disorder. Instead, we provide the sensitivity and specificity for different cut off scores based on our small group of 12 deaf individuals with dementia in table 9 to assist clinicians in carefully evaluating the likelihood of dementia. These data show, for example, that a cut-off of 14 would correctly classify 83.3% (n=10) of dementia patients and misclassify 7.6% (n=17) of controls as belonging to the dementia group. A lower cut-off point would increase false negatives, wrongly classifying dementia patients as being cognitive healthy; and a higher cut-off point would increase false labeling of controls as having dementia.

False positive recognition errors

The sensitivity and specificity of different cut offs are shown on table 10. A comparison of the percentage of patients and controls making false positive recognition errors are shown in Table 11. The large majority of healthy controls (85.9%) made no errors and only 4.8% made two or more errors. The majority of patients with dementia (64.7%) made at least one error, with 41.2% making two or more. These figures suggest that individuals making 2 or more false positive errors have a greater likelihood of belonging to the dementia group and accuracy of correct classification increases with a greater number of errors.

Discrimination index

The discrimination index measuring the difference between true positive and false positive recognition responses was at ceiling for the 70.9% of controls that made zero errors. Discrimination scores for dementia patients ranged from 0-11, with half obtaining <7 compared to only 1.3% of controls. Table 12 provides the sensitivity and specificity of different cut off scores.

Discussion

The aim of this study was to furnish practitioners with useful normative data to enable the assessment of verbal memory impairment in deaf people who use BSL. The BSL-VLMT is the only verbal memory test developed for a signed language with norms for older adults. Our data show that it can be successfully administered to both cognitively healthy adults and adults with dementia, with excellent interrater reliability.

Our patient comparison study was explorative and small scale and conclusions must be tentative due to small sample size. Obtaining a large clinical sample of deaf people with a confirmed dementia is very challenging. Deaf people are a minority group, and there were no existing standard measures for diagnosing dementia in this population. The paper presents preliminary data from 12 cases. These suggest that BSL-VLMT has promising clinical utility for detecting dementia. Although cut off scores are identified, they should be applied cautiously as one potential indicant of a diagnosis. Further research with a larger clinical sample, different aetiologies, those with MCI, and the collation of norms for younger adults, and consideration of dementia severity, would provide more thorough validation for clinical use.

A weakness of this study is that it does not question the underlying construct of verbal memory in deaf signers and how it may differ from users of spoken languages who can hear. Experimental studies have found differences in the verbal short term memory of deaf people, which may be due to the nature of sign language and/or different neural organization (Emmorey & Wilson, 2004; Rudner, Andin, & Ronnberg, 2009; Wilson & Emmorey, 1997). It is important for researchers who might want to use these norms to consider differences in articulation rates and phonological similarity for signs versus spoken words which may influence memory capacity, encoding and retrieval (see Wilson & Emmorey, 2000 for a review), and normative differences in memory for serially presented material in deaf people (Boutla, Supalla, Newport, & Bavelier, 2004), such as the tendency to show less temporal order effect in free recall tasks (Bavelier et al., 2008).

The influence of demographic variables on BSL-VLMT scores mirrors spoken language studies using the HVLT-R. Performance decreases with age (Benedict et al., 1998; Vanderploeg et al., 2000) and there is a small gender effect with females outperforming males on immediate recall (Vanderploeg et al., 2000; Friedman et al., 2002). Some HVLT studies found an effect of length of education (Friedman et al., 2002; Hester et al., 2004) but others reported no effect (Vanderploeg et al., 2000). Years of education had no influence on BSL-VLMT scores once age was taken into account. This variable may be problematic in relation to the Deaf population because it taps into historical changes in deaf education relating to language of instruction, access to the curriculum and quality of teaching, which has often been influenced by educational policy rather than providing a measure of ability.

The absence of comparator tests also makes it difficult to establish construct validity. Scores on the BSL-VLMT were compared to the only other test of memory developed for users of BSL, with a significant but moderate correlation. This probably reflects the fact that while both tasks assess memory, they impose different demands; i.e. one required recollection of a list of words and the other facts about a described individual. No gold standard contrastive test enabled us to assess convergent validity. Likewise further work is required to ensure discriminant (or divergent) validity for deaf signers using other measures once they are developed for this population.

We recommend that clinicians administering the test should be proficient users of British Sign Language to enable effective clarification and repetition of instructions, accurate recording and scoring. Administration should be possible via a professionally qualified interpreter with careful preparation and briefing, but this has not been empirically validated. We also recommend that patient responses be filmed for later verification.

Although the current sample included people who became deaf between birth and 10 years old, with different aetiologies of deafness and ages of BSL acquisition, these variables did not show a relationship with verbal memory or learning. The BSL-VLMT is untimed and involves simple lexical list learning and does not require the maintenance of temporal order. Our study shows that this paradigm has validity as a measure across a heterogeneous deaf population. Differences in language ability were related to verbal memory performance but not to the ability to learn across trials. It is well established that language ability correlates with short-term verbal memory (e.g. Cantor, Engle, & Hamilton, 1991) so it is not surprising that we have found the same relationship in deaf signers. An important omission from this study is that we did not collect information about degree of bilingualism which is relevant because research shows that deaf people that learn sign language as an L2 can achieve near native levels of proficiency providing they have a strong L1 (Cormier, Schembri, Vinson, & Orfanidou, 2012; Mayberry, 2007).

A third of our normative sample were unable to produce any items at delayed recall despite the majority having good performance on immediate recall, learning and recognition. This pattern differs from individuals with Alzheimer's disease dementia who typically show poor learning and encoding; with rapid forgetting and inability both in delayed recall and recognition (DeFina, Moser, Glenn, Lichtenstein & Fellus, 2013). It was our observation during normative data collection that most zero scores were caused by interference effects due to participants not understanding which of the several tests in the test battery were referred to in the instructions for delayed recall. It was not that participants were unable to remember any items, but that they did not understand which test items they were being asked to recall. The finding of lower language and nonverbal intellectual ability among zero-scorers supports this explanation. Caution is therefore required in interpreting a delayed recall score of zero which is not necessarily pathological and should be considered relative to the other BSL-VLMT scores. Where it is combined with low percentile score on immediate recall, poor learning over trials and/or recognition errors there will be cause for clinical concern. For clinical use, we recommend checking for genuine delayed recall impairment by providing respondents who score zero with a cue for retrieval (i.e. the first item in the recall list: MOUSE). We did not collect cue data from our normative group, however previous research shows that prompts substantially improved recall in healthy individuals (e.g. Ivanoui et al., 2005) whereas our patients with dementia showed no improvement with cueing (see also Davis & Mumford, 1984).

It is important to acknowledge the influence of sampling on this study. We chose not to exclude outliers so that the norms capture the range of ability within the Deaf population, which is heterogeneous in nature, with different ages of language acquisition, language fluency and deafness onset. The lack of accurate assessment tools and provision means that hidden cognitive impairments (either developmental or acquired) may be overlooked in this population. Our research group used an almost identical normative sample for the development of a BSL Cognitive Screening Test and reported that despite pre-screening the sample contained cases of undiagnosed learning disability or acquired cognitive impairment, with a hidden, low ability bump in the bimodal distribution of nonverbal intellectual ability for those in their 50s and 60s, which may reflect a low ability bias among the working age people who chose to attend holiday camps for older adults, and also indicates the need for better services and diagnostic tests (Atkinson et al., 2015). The current normative sample contained 10 individuals with *borderline* to *severely impaired* nonverbal intellectual ability; of these 6 were also BSL-VLMT outliers with either low performance on learning trials or recognition errors or both. Two had a very low score on the BSL Cognitive Screening Test that was suggestive of dementia rather than developmental impairment. A further 6 people with normal intellectual ability had low immediate recall scores, with 3 also making recognition errors, but with a BSL-CST score above the level indicative of dementia. This suggests the current test was detecting cases of

MCI in our normative sample. Rather than excluding these outliers, they were retained in the norms to ensure the sample represents the full spectrum of those living within the Deaf Community. The scores fall in the lowest percentiles, which reflect the conventional levels for clinical concern/impairment.

Late diagnosis and more severe levels of impairment in our dementia sample might explain why our measures showed such an excellent ability to correctly distinguish patients with a confirmed diagnosis of dementia from controls. If our patient sample had included more cases of MCI we might have found a lower rate of diagnostic accuracy due to greater overlap between our two samples.

A limitation of the BSL-VLMT is the existence of only one version of the test, unlike the HVLT-R, which has six different forms designed to reduce practice-related error for patients undergoing serial testing. We were constrained by the limited pool of BSL signs with data about lexical properties from which to draw our stimuli. This underlines the importance of lexical norming work for signed languages to provide a resource for further test development.

The format and method reported here are transferable to sign languages in other countries, although direct translation will not work because of differences in lexical variables and cultural familiarity. Unique test stimuli would be needed for each national sign language, carefully selected according to lexical ratings for that language, and new normative data would need to be collected from the local Deaf population. Some of the stimuli may be transferrable to other signed languages in the BSL family including Australian Sign Language (AUSLAN) and New Zealand Sign

Language, which are similar enough for mutual understanding of some lexical items. Stimuli which differ could be replaced with local signs with careful attention paid to the neuropsychological implications of such modification. The development of this test and its potential for adaptation into other sign languages is a significant step towards more equitable assessment and timely diagnosis for deaf users of sign language who present with memory disorders.

References

Bavelier, D., Newman, A. J., Mukherjee, M., Hauser, P., Kemeny, S., Braun, A., & Boutla, M. (2008). Encoding, rehearsal, and recall in signers and speakers: shared network but differential engagement. *Cereb Cortex*, *18*(10), 2263–2274.

Boutla, M., Supalla, T., Newport, E. L., & Bavelier, D. (2004). Short-term memory span: insights from sign language. *Nature Neuroscience*, *7*(9), 997–1002. doi:10.1038/nn1298

Cantor, J., Engle, R. W., & Hamilton, G. (1991). Short-term memory, working memory, and verbal abilities: How do they relate? *Intelligence*, *15*(2), 229–246. doi:10.1016/0160-2896(91)90032-9

Cormier, K., Schembri, A., Vinson, D., & Orfanidou, E. (2012). First language acquisition differs from second language acquisition in prelingually deaf signers: Evidence from sensitivity to grammaticality judgement in British Sign Language. *Cognition*, *124*(1), 50–65. doi:http://dx.doi.org/10.1016/j.cognition.2012.04.003

Dubois, B., Feldman, H. H., Jacova, C., Dekosky, S. T., Barberger-Gateau, P., Cummings, J., … Scheltens, P. (2007). Research criteria for the diagnosis of

Alzheimer’s disease: revising the NINCDS-ADRDA criteria. *Lancet Neurology*, *6*(8), 734–46. doi:10.1016/S1474-4422(07)70178-3

Emmorey, K., & Wilson, M. (2004). The puzzle of working memory for sign language. *Trends Cogn Sci*, *8*(12), 521–523.

Mayberry, R. I. (2007). When timing is everything: Age of first-language acquisition effects on second-language learning. *Applied Psycholinguistics*, *28*(03), 537–549. doi:10.1017/S0142716407070294

McKhann, G., Drachman, D., Folstein, M., Katzman, R., Price, D., & Stadlan, E. M. (1984). Clinical diagnosis of Alzheimer’s disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology*, *34*(7), 939–44. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/6610841

Morere, D. a. (2013). The Signed Verbal Learning Test: Assessing Verbal Memory of Deaf Signers. *Sign Language Studies*, *14*(1), 39–57. doi:10.1353/sls.2013.0025

Neary, D., Snowden, J. S., Gustafson, L., Passant, U., Stuss, D., Black, S., … Benson, D. F. (1998). Frontotemporal lobar degeneration: a consensus on clinical diagnostic criteria. *Neurology*, *51*(6), 1546–54. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9855500

Rudner, M., Andin, J., & Ronnberg, J. (2009). Working memory, deafness and sign language. *Scandinavian Journal of Psychology*, *50*(5), 495–505. doi:10.1111/j.1467-9450.2009.00744.x

Wilson, M., & Emmorey, K. (1997). A visuospatial “phonological loop” in working memory: Evidence from American Sign Languageitle. *Memory & Cognition*, *25*(3), 313–320.

Wilson, M., & Emmorey, K. (2000). When does modality matter? Evidence from ASL on the nature of working memory. In K. Emmorey & H. Lane (Eds.), *The signs of language revisited. An anthology to honor Ursula Bellugi and Edward Klima* (pp. 135–142). Mahwah, NJ: Lawrence Erlbaum Associates.

Zhou, X., & Obuchowski, N Obushcowski, D. (2002). *Statistical methods in diagnostic medicine*. New York: Wiley & Sons.

Table 1: Demographics for control and dementia groups

	Control N=223			**Dementia N=12**			**Pearson or Mann Whitney U**
	M	**SD**	**Min-Max**	**M**	**SD**	**Min-Max**	
Age	68.13	10.05	50-89	75.77	9.25	54-88	†
Years of education	11.39	2.09	4-20.5	10.57	4.90	6-32	
Age of BSL acquisition	6.45	5.13	0-40	4.85	3.24	0-11	

†$p<0.001$ (two-tailed)

Table 2: Demographic variables, mean BSL-VLMT total immediate recall scores (sum of trials 1-3), standard deviations and statistical values

Variable	**N=223**	**M (SD)**	**r/r_s /F/t**
Age		21.71 (5.11)	r(221)=-0.366 p=<0.001* d=-0.787
50-59	51	24.63 (4.27)	F(3,219)=10.39 p<0.001* d=0.124
60-69	76	21.87 (4.69)	
70-79	50	20.80 (4.88)	
80-89	46	19.67 (4.59)	
Gender			
Female	143	22.50 (4.77)	t(221)=-2.879 p<0.01* d=0.399
Male	80	20.56 (4.93)	
Nonverbal ability			t(221)=.423 p<0.001*
Region			
South East England	102	21.05 (5.15)	t(221)=2.135 p<0.05* d=0.286
Other UK regions	121	22.45 (4.62)	
Age became deaf			r_s(221)=-0.045 p=0.507
Age of BSL acquisition #			r(221)=-0.083 p=0.219
Native	26	23.62 (5.55)	F(2,218)=2.084 p=0.127
Early (1-5 yrs)	92	21.61 (4.91)	
Late (6+ yrs)	103	21.48 (4.70)	
Cause of deafness #			
Genetic	69	22.82 (4.79)	F(2,216)=2.520 p=0.083
Organic	49	21.47 (4.30)	
Other/unknown	101	21.15 (5.20)	
Years of education (from 5 years) #			r(221)=0.172 p<0.05* d=0.349
7-9	32	20.22 (5.44)	F(4, 216)=2.108 p=0.081
10-12	146	21.92 (4.80)	
13-14	24	21.25 (5.06)	
15-16	14	24.21 (4.17)	
17-21	5	24.20 (3.27)	
Occupational status			

Professional	7	23.57 (4.39)	F(4, 218)= 0.847 p=0.497
Intermediate	7	24.00 (2.52)	
Skilled	77	22.08 (4.44)	
Semi-skilled	76	21.57 (5.83)	
Unskilled	56	21.27 (4.39)	
Educational attainment			
Degree/postgraduate	11	24.72 (4.92)	F(5,217)=2.953 p<0.05* d=0.064
A level or equivalent	4	25.50 (4.65)	
O level/CSE/GCSE or	13	24.62 (4.03)	
BSL teaching	13	23.31 (3.61)	
Vocational	43	21.44 (5.67)	
None	139	21.18 (4.66)	

*Asterisk marks significant result

# Hashtag indicates missing data: Age of BSL acquisition 2 cases, Cause of deafness 4 cases, Years of education 2 cases

Table 3: Percentile scores for immediate recall (sum of trials 1-3)

	Percentile scores						
	1st-2nd	5th	10th	25th	50th	75th	90th
50-59	10.2	16.2	19.2	22	25	28	29
60-69	*10.1*	13.6	16.2	20	23	25	27
70-79	10.04	12	14.1	17	21	25	*26*
80-89	7	9	10.8	*16*	20	23	25.2

*Italics: scores smoothed to take account of sampling effects and irregularities (Rust & Golombok, 1999)

n= 223

Table 4: Percentile cut offs for learning index

Learning index	
1	<5th
2	10th
3	25th
4	50th
5	75th
6	80th
7	>90th

n=223

Table 5: Percentile scores for delayed recall for participants scoring greater than zero (n=144)

	Percentile scores					
	<5th	10th	25th	50th	75th -90th	>90th
50-59	6	7	9	10	11	12
60-69	6	7	7	10	11	12
70-79	6	7	8	9	9	11
80-89	2	5	6	8	9	10

Table 6: Percentile cut offs for recognition

Recognition score (true positives)		False positive errors		Discrimination index	
10 or less	<5th	1	>10th	2 or less	<1st
11	10th	2 or more	<5th	6	2nd
12	>25th			8	5th
				10	10th
				11	25th
				12	>30th

n= 223

Table 7: Comparison of control and dementia groups aged 70-89 years on demographics and BSL-VLMT scores

	Control N= 95			**Dementia N= 10**			
	M	**SD**	**Min-Max**	**M**	**SD**	**Min-Max**	**Mann Whitney U**
Age	77.75	5.66	*70-89*	79.70	5.18	73-88	
Years of education	10.6	1.50	7-15.5	10.45	1.11	8-11.5	
Age of BSL acquisition	6.72	5.17	0-40	4.70	2.67	0-9	
Immediate recall (sum of trials 1-3)	20.27	4.78	7-31	11.1	6.31	1-20	†
Learning index	4.21	1.72	-1-8	2.00	1.33	0-4	†
▪ Intrusions*	0.73	1.35	0-8	2.90	3.14	0-8	†
▪ Repetitions *	0.75	1.29	0-7	1.30	3.43	0-11	
▪ Clusters*	4.08	1.68	0-9	1.80	1.93	0-6	#
Delayed recall	4.25	4.37	0-12	1.00	3.16	0-10	‡
Retention index	46.60	46.60	0-133.33	12.5	39.53	0-125	‡
Recognition (true positive score)	11.63	0.80	8-12	10.70	2.36	5-12	
False positive errors	0.26	0.67	0-4	2.3	2.06	0-6	†
Discrimination index	11.30	1.33	6-12	7.00	3.80	0-11	†

†*p*<0.001, ‡*p*<0.01, *#p<0.05* **(two-tailed)**
*** immediate recall trials (1-3)**

Table 8: Areas under curves for BSL-VLMT measures

BSL-VLMT Measure	**Area under curve (AUC)**	**Standard error**	**P**	**Sensitivity and specificity trade-off**
Immediate recall (sum of trials 1-3)	.945	.026	<.001	Excellent
Learning index	.829	.055	<.001	Good
Delayed recall	.823	.036	<.001.	Good
Retention index	.829	.055	<.001	Good
False positive errors	.913	.049	<.001	Excellent
Discrimination index	.950	.019	<.001	Excellent

Table 9: Sensitivity-specificity trade-offs of different cut scores for immediate recall (sum of trials 1-3)

Immediate recall score	**Sensitivity**	**Specificity**
<10	0.500	0.964
<11	0.583	0.964
<12	0.750	0.946
<13	0.750	0.937
<14	0.833	0.924
<15	0.833	0.892
<16	0.917	0.865
<17	0.917	0.825
<18	0.917	0.780
<19	0.917	0.722
<20	1.00	0.650
<21	1.00	0.543

Table 10: Sensitivity-specificity trade-offs of different cut scores for number of false positive recognition errors

False positive errors	Sensitivity	Specificity
0	1.000	0.000
<1	0.583	0.951
<2	0.500	0.969
<3	0.250	0.991
<5	0.167	1.000

Table 11: Percentage of participants making false positive errors on recognition task

False positive errors	% Control	% Dementia
>0	100	100
>1	14.1	64.7
>2	4.8	41.2
>5	0.8	17.7

Table 12: Sensitivity-specificity trade-offs of different cut off scores for discrimination index

False positive errors	Sensitivity	Specificity
<12	1.000	0.000
<11	1.000	0.709
<10	0.917	0.892
<9	0.833	0.933
<8	0.667	0.955
<7	0.500	0.969