

TITLE PAGE

Title

A valid alternative for in-person language assessments in brain tumor patients: Feasibility and validity measures of the new TeleLanguage Test

Running title

A Telephone-based Language Test for patients

Elke De Witte, Vitória Piai, Garret Kurteff, Ruofan Cai, Peter Mariën[†], Nina Dronkers, Edward F. Chang, Mitchel S. Berger

Affiliations:

University of California San Francisco Medical Center, Department of Neurological Surgery, San Francisco, California [EDW, GK, EC, MB]

Vrije Universiteit Brussel, Department of Clinical and Experimental Neurolinguistics, Center of Linguistics and Literary Studies, Brussels, Belgium [EDW, PM]

Radboud University Medical Center, Department of Medical Psychology, Nijmegen, the Netherlands [VP]

ZNA Middelheim, Department of Neurology & Memory Clinic, Antwerp, Belgium [PM]

Radboud University, Donders Centre for Cognition, Nijmegen, the Netherlands [VP]

VA Northern California Health Care System, Center for Aphasia and Related Disorders, Martinez, California [ND]

Corresponding author

Elke De Witte

Vrije Universiteit Brussel

Department of Clinical and Experimental Neurolinguistics

Center of Linguistics and Literary Studies (LIST)

Pleinlaan 2, 1050 Brussels, Belgium

elkedewitte@gmail.com, +1 415 518 3352

Funding

Elke De Witte is a postdoctoral fellow of the Research Foundation - Flanders (FWO).

VP was funded by the Netherlands Organization for Scientific Research (grant number 446-13-009).

Conflict of interest

There is no conflict of interest.

- Mention of total manuscript word count: within 6000 words (excl tables)

Abstract

Background: Although language deficits after awake brain surgery are usually milder than post stroke, postoperative language assessments are needed to identify these. Follow-up of brain tumor patients in certain geographical regions can be difficult when most patients are not local and come from afar. We developed a short telephone-based test for pre- and postoperative language assessments.

Methods: The development of the TeleLanguage Test was based on the Dutch Linguistic Intraoperative Protocol and existing standardized English batteries. Two parallel versions were composed and tested in healthy native English speakers. Subsequently, the TeleLanguage Test was administered in a group of 14 tumor patients before surgery and at 1 week, 1 month and 3 months after surgery. The test includes auditory comprehension, repetition, semantic selection, sentence/story completion, verbal naming and fluency tests. It takes less than 20 minutes to administer.

Results: Healthy participants had no difficulty performing any of the language tests via the phone, attesting to the feasibility of a phone assessment. In the patient group, all TeleLanguage test scores significantly declined shortly after surgery with a recovery to preoperative levels at 3 months postsurgery for naming and fluency tasks and a recovery to normal levels for the other language tasks. Analysis of the in-person language assessments (until 1 month) revealed a similar profile.

Conclusion: The use of the TeleLanguage battery to conduct language assessments

from afar can provide convenience, might optimize patient care and enables longitudinal clinical research. The TeleLanguage is a valid tool for various clinical and scientific purposes.

Key-words: tumor surgery, language test, telephone assessment, awake brain surgery, aphasia

Importance of the study:

Follow-up of brain tumor patients undergoing surgery can be difficult when patients are not local and come from afar. Therefore, we developed a valid alternative for in-person language assessments in brain tumor patients: the TeleLanguage Test for pre- and postoperative language assessments. No other language telephone test batteries exist for this purpose. The administration of the TeleLanguage Test in both healthy participants and patients showed that feasibility and preliminary validity of this test is high. Therefore, the TeleLanguage Test is a valid tool for various clinical and scientific purposes in the brain tumor group. In addition, it could be tested in other patient populations where there is a high burden associated with frequent follow-up in-person testing leading to drop-outs (e.g. epilepsy patients).

INTRODUCTION

Awake brain surgery with intraoperative mapping is currently the gold standard method for brain tumors (gliomas) located in functional regions ¹ which typically involve language or motor cortices/pathways of the dominant hemisphere ². According to the meta-analysis of De Witt Hamer et al. ³ intraoperative mapping with Direct Electrical Stimulation (DES) is associated with fewer late severe neurological deficits compared to glioma surgery without DES. Still, a number of neuropsychological studies have shown that both in the pre- and postoperative phase of awake surgery relevant clinical language impairments may be found ⁴⁻⁹ and patients may suffer from aphasia, i.e., inability to understand or formulate language, as a result of brain injury. However, due to the typically slow growth of low-grade gliomas allowing neural reorganization, no obvious linguistic deficits may be identified when using standardized language tests for stroke patients ^{1,10}. Therefore, more sensitive language tests are needed to identify the more subtle language impairments tumor patients complain about. Consequently, the Dutch Linguistic Intraoperative Protocol (DuLIP) ¹¹ was developed to measure language functions in the pre-, intra- and postoperative phase of awake surgery. This protocol consists of tasks and items at all linguistic levels, i.e. phonology, semantics, and (morpho)syntax at different difficulty levels for language production and perception.

Wilson et al. ¹² evaluated the language of 110 patients from the University of California San Francisco (UCSF) Medical Center. The patients were assessed prior to surgery and 2-3 days and 1 month postsurgery using the Western Aphasia Battery (WAB) ¹³ and the Boston Naming Test ¹⁴. Their findings showed that transient aphasia is very common (in 71% of patients) after left hemisphere surgery and that the nature

of the aphasia depends on the site of the specific location of the surgical site. At 1 month postsurgery, only naming differed significantly from its presurgical level. More recently, a newly composed test battery, the quick aphasia battery (QAB)¹⁵ is applied at UCSF for pre- and postoperative assessments of brain tumor patients. The QAB was developed to gather relevant information in a short period of time on the different linguistic levels and major language functions in patients. In addition, as the QAB uses graded systems to quantify deficits in patients it might be more sensitive than the WAB to detect mild language deficits.

Unfortunately, postoperative language follow-up of tumor patients in the San Francisco Bay Area can be difficult as many are referred from outside the area and find it difficult to return for follow-up testing. Many other US and European sites experience the same problem, particularly with extended follow-ups. Consequently, few long-term data are collected even though it has been shown that language recovery can take longer than several months^{4-6,16-19}. Administering language tests such as the WAB^{13,14} or QAB via video calling (e.g., Skype) has proved inefficient as patients are not always able to use video calling and/or their internet connection could not support it. Since most tumor patients are typically contacted by phone post-surgery for clinical follow-up, it seems more feasible to do follow-up language testing via the telephone rather than via video calling. In fact, numerous studies have shown that various cognitive functions can be measured reliably and precisely over the telephone in stroke, cancer, Alzheimer's disease and the elderly²⁰⁻²⁴. However, no language-focused telephone assessments are available, nor have existing standardized language batteries been tested via the telephone. To fill this void, we developed a short telephone-based language test battery for pre- and postoperative language assessments in order to follow patients over longer periods of time. In this pilot study, a group of 14 brain tumor patients undergoing awake

surgery for resection of language dominant hemisphere tumors and 14 matched controls were evaluated with a telephone-based test battery. The objectives of this study were: 1) to compose a feasible telephone-based language test battery for brain tumor patients, 2) to collect longer follow-up data (currently at 3 months post-surgery), 3) to validate the telephone-based language battery by comparing the collected telephone data with traditional face-to-face assessments.

MATERIALS AND METHODS

Development of the TeleLanguage Test

Procedure

Task selection for the TeleLanguage Test was based on the Dutch Linguistic Intraoperative Protocol ¹¹ and English aphasia batteries such as the WAB ¹³, Boston Diagnostic Aphasia Examination ²⁵ and the QAB ¹⁵. We composed a language battery consisting of a naming task, phonological tasks (repetition, letter fluency), semantic tasks (semantic selection, semantic fluency) and syntactic language tests (story completion, sentence completion) using different difficulty levels in order to identify both mild and more severe language deficits. A comprehension screening was added to exclude patients with severe comprehension disorders who would not be able to perform the tests via the telephone. The test items were selected on the basis of linguistic variables, such as word frequency, word length, morphological and phonological form such that each task had both easy and more complex items, thus with increasing difficulty level. We developed two parallel versions of the TeleLanguage Test (version A and B, except for the fluency tasks) that were matched to avoid learning/practice effects. These 2 versions were tested in a group of 10 healthy native English speakers to evaluate the selection of items. The administration of the whole TeleLanguage Test did not last longer than 20 minutes. Based on the assessment in the healthy test group, some items from the semantic selection task were adapted (e.g. when the target answer was not found by more than 2 subjects) and some repetition items (including many high frequency sounds) were deleted to avoid misunderstanding via the telephone before administration in the tumor group.

Description of tests

All tests used are illustrated in Table 1, including their provenance and number of items.

Test administration, registration and scoring

In each version (A or B), the tests were administered in identical order: 1. comprehension screening, 2. verbal naming, 3. repetition, 4. semantic selection, 5. story completion, 6. letter fluency, 7. semantic fluency, 8. sentence completion. Practice items were included for each subtest to ensure that participants understood the task. The answers were recorded verbatim and transcribed. One point was given if the target answer was correctly produced and within the given time frame for the fluency and naming task (within 10 seconds). For the repetition task, minor dysarthric errors are allowed and articulation errors were marked in the column response. A single repetition by the examiner may be given if requested (e.g. if the patient asks or does not seem to hear). The accuracy of the scoring of the tests was checked by the first author and was based on qualitative terms agreed on a priori between the first and second authors. The administrators were speech language pathologists and neuropsychologists trained in language assessments. Uncertainties regarding the correctness of test items were discussed with the other authors until a consensus was reached.

Task	Example (Item)
-------------	-----------------------

Comprehension screening (n = 5; Adapted from WAB)	I am going to ask you some questions. Answer yes or no: 1. Is your name X? 2. Do you eat a banana before you peel it?
Verbal naming test (n = 25) (Adapted from verbal naming test; ²⁶)	I am going to describe an object or a verb and I want you to tell me the name of what I am describing: 1. A large animal in Africa with a trunk? 2. What ice does when it gets hot?
Word & sentence repetition (n = 10; Adapted from WAB, BDAE)	Repeat after me: 1. Bed 2. Screwdriver 3. Methodist episcopal 4. He unlocks the heavy oak door.
Letter fluency (Adapted from WAB)	Name as many words that start with the letter F as you can in one minute.
Semantic noun & verb selection (n = 10; Adapted from DuLIP)	Two words will go together best because of their meaning and one word will not. Tell me the word that does not fit? 1. Banana, apple, carrot 2. Talk, tell, sing
Semantic fluency (Adapted from WAB)	Name as many animals as you can in one minute.

<p>Story completion (n = 5; Adapted from Goodglass story completion ²⁷)</p>	<p>Complete the story:</p> <ol style="list-style-type: none"> 1. My dog is hungry and I have a bone in my hand. What's next? 2. The mouse was running around. A cat came along. The mouse did not see the cat running after him. What happened to the mouse? The mouse ...
<p>Sentence completion (n = 10; Adapted from DuLIP)</p>	<p>Complete the sentence in a meaningful way</p> <ol style="list-style-type: none"> 1. The man knows that ... 2. I'll do it when ...

Table 1: Description of TeleLanguage tasks with item examples

Clinical use of TeleLanguage in tumor patients and matched healthy controls.

Participants

A consecutive series of 18 patients with left hemisphere tumors, who all underwent intraoperative language mapping during resection, were screened for this study. Exclusion criteria included: non-fluent English speakers, a history of a medical or psychiatric condition known to affect language functions, persistent language deficits as a result of prior treatment, mild to severe preoperative language deficits (less than 4 out of 5 for the comprehension screening), auditory or severe visual disorder, and mental retardation. Four patients were excluded because of non-native/non-fluent English, consequently 14 patients were administered the TeleLanguage protocol (Table 2 for demographics and clinical characteristics). Handedness was formally assessed by the short form of the Edinburgh Handedness Inventory ²⁸ which revealed 14 right-handed patients. Intraoperative and postoperative reports revealed detection of cortical

language sites in all but three patients (Patients 2, 3, and 11), detection of subcortical language sites in Patients 3, 4, 7 and 14, and concomitant apraxia of speech in Patients 3, 7 and 14.

The 14 healthy participants were matched for gender (four females, ten males), age (mean: 55.29 years, SD: 10.33, range 39-71) and education levels (mean: 15.64 years, SD: 2.24, range 12-19). Exclusion criteria included: non-fluent English speakers, history of psychiatric diseases, alcohol and/or drug addiction, sleep medication, impaired hearing or vision. Informed consent for the study was obtained from all participants. This study was performed under a protocol approved by the UCSF Committee on Human Research.

Pt	Gender	Age	Education	Handedness	Tumor location	Tumor type & grade
Patients with highest mean language scores						
2	M	63	12	Right	L frontal	Glioblastoma IV
5	M	63	15	Right	L temporal	Glioblastoma IV
10	M	52	18	Right	L temporal	Glioblastoma IV
11	M	52	16	Right	L frontal	Oligodendroglioma III
Patients with lowest mean language scores						
3	M	50	15	Right	L temporal/insular	Glioblastoma IV
4	F	67	12	Right	L parietal	Glioblastoma IV

7	M	58	14	Right	L temporal/insular	Oligodendroglioma II
14	M	39	17	Right	L temporal	Glioblastoma IV
Patients with middle mean language scores						
1	F	50	14	Right	L temporal	Oligodendroglioma II
6	F	68	13	Right	L parietal	Glioblastoma IV
8	M	52	19	Right	L insular	Glioblastoma IV
9	M	42	17	Right	L frontal	Oligodendroglioma II
12	M	66	19	Right	L frontal	Oligodendroglioma II
13	F	50	16	Right	L temporal	Oligodendroglioma II

Table 2: Demographics and clinical characteristics of patients; M = male, F = female, L = left; a mean language score was calculated per patient by pooling all postoperative scores per subtest together and taking the mean of these TeleLanguage test scores (using percentages), see Results section for clarification.

Procedure

The TeleLanguage Test was administered to 14 tumor patients prior to surgery and again within 1 week, 1 month, and 3 months after surgery. Seven patients received the protocol in A-B-A-B order (Patients 1,3,5,7,9,11,13) and seven patients (2,4,6,8,10,12,14) in the reverse order, i.e. B-A-B-A.

When possible, in-person assessments were also performed in this patient group in order to compare TeleLanguage data with data collected face-to-face. The in-person assessments consisted of the QAB (3 parallel versions) prior to surgery, within one week after surgery, and at 1 month after surgery. A selection of subtests from the QAB

was used to compare with TeleLanguage (Table 3 for the comparison). We note that tests do not necessarily have a one-to-one mapping between TeleLanguage and the QAB, but the best possible analogues were used.

In order to investigate how the patients scored in comparison to their healthy peers, 14 matched controls (according to gender, age, education) were selected to participate in this study and also underwent the TeleLanguage protocol. Both version A and B were administered with two weeks in between: seven controls (uneven numbers) in A-B order, seven controls (even numbers) in B-A order.

PHONE ASSESSMENTS (TeleLanguage)	IN-PERSON ASSESSMENTS (QAB)
Verbal naming	QAB naming: visual naming of 6 colored pictures (n = 6)
Repetition	QAB repetition: repetition of words (n = 4) and sentences (n = 2) increasing in length
Semantic selection	QAB semantic non-verbal cognition: Participants were asked to find the picture (1 out of 6) that did not fit on the basis of semantic relationships (only the first 2 series were selected because the other 4 series of odd picture out were based on visual categorization instead of semantic categorization as in the phone test) (n = 2)

<p>Story completion with full sentence</p>	<p>QAB sentence production: The subject needs to describe two stimulus cards. 'What is happening here?' e.g. The boy is pushing the girl. (n = 2)</p>
<p>Sentence completion</p>	<p>QAB connected speech A conversation with the subject for at least 3 minutes, around one or more topics. The following scales were rated on a 0 (mute) to 6 score (normal): paragrammatism, agrammatism, semantic paraphasia, phonemic paraphasia, other lexical access difficulties. These scales were selected to compare with sentence completion from TeleLanguage.</p>

Table 3: Selection and description of subtests from the QAB that are used to compare with the subtests from the TeleLanguage Test.

Statistics

First, the non-parametric paired Wilcoxon signed-rank test was used to compare the healthy control data from version A with those from version B of the TeleLanguage Test, in order to define whether version A and B were indeed similar. Second, we investigated if the mean pre- (T1) and postoperative raw TeleLanguage test scores (T2, T3, T4) of the patients differed significantly from the healthy controls using a Mann-Whitney U test. Third, the patients' pre- and postoperative raw scores were compared for both the TeleLanguage assessments and the in-person assessments using non-parametric paired Wilcoxon signed-rank tests. For the TeleLanguage assessments, z

scores were also calculated (using the mean and standard deviation from the healthy control data) to identify whether a score was clinically (≤ -1.5 SD) or pathologically (≤ -2 SD) impaired²⁹. Finally, given that normative data for the QAB subtests are not yet available, the in-person data and TeleLanguage data were transformed into percentages (i.e., the raw score divided by the maximum score and multiplied by 100) and compared with a paired Wilcoxon signed-rank test.

Analyses were performed using IBM SPSS Statistics 23 and results were considered significant at a $p < .05$ level. As this is a small exploratory study, we did not correct for multiple comparisons.

RESULTS

Healthy controls

Maximum scores for all healthy controls were observed for the comprehension screening (max 5), repetition task (max 10), story completion task (max 5) and sentence completion task (max 10). For naming and semantic selection mean scores of 24.71 (version A: 24.74, version B: 24.67, max 25) and 9.63 (version A: 9.5, version B: 9.75, max 10) were respectively found. Only 1 or 2 errors were made by a small number of healthy controls for naming and semantic selection.

In the healthy control group there was no significant difference in performance between the language tasks from version A and B (for all language tasks: $p \geq .257$).

Patients vs healthy controls

Patients' preoperative scores (T1) were worse compared to healthy controls' on the verbal naming task ($p = .001$), letter fluency ($p = .008$) and semantic fluency task ($p = .001$) (see Table 4 for patient's raw scores).

At 1 week postsurgery (T2), patients obtained significantly worse scores for all language tasks: verbal naming ($p < .0001$), repetition ($p = .039$), semantic selection ($p < .0001$), story completion ($p = .014$), letter fluency ($p < .0001$), semantic fluency task ($p < .0001$) and sentence completion ($p < .0001$).

At 1 month postsurgery (T3), almost the same psychometric profile was found as at T2. In contrast with T2, only semantic selection ($p = .578$) recovered to the normal range of healthy controls (verbal naming: $p = .001$, repetition: $p = .016$, story completion: $p = .016$, letter fluency: $p = .001$, semantic fluency task: $p < .0001$, sentence completion: $p = .0037$).

At 3 months postsurgery (T4), performance on verbal naming ($p = .045$), letter fluency ($p = .05$) and semantic fluency ($p = .01$) was still significantly lower than for healthy controls.

Within patients, between test moments

Using raw scores ($N = 14$), the TeleLanguage assessments revealed a significant decline at 1 week postsurgery for all language tasks: verbal naming ($p = .003$), repetition ($p = .026$), semantic selection ($p = .003$), story completion ($p = .010$), letter fluency ($p = .002$), semantic fluency ($p = .002$) and sentence completion ($p = .002$). During short-term follow-up (between 1 week and 1 month postsurgery) a significant improvement was observed for verbal naming ($p = .002$) and semantic selection ($p = .007$). During long-term follow-up (between 1 week and 3 months postsurgery) a significant improvement was observed for all language tasks: verbal naming ($p = .003$), repetition ($p = .027$), semantic selection ($p = .005$), story completion ($p = .013$), letter fluency ($p = .002$), semantic fluency ($p = .002$) and sentence completion ($p = .002$). Semantic fluency ($p = .005$) was the only task that still showed impairment at 1 month postsurgery relative to the preoperative assessments. No significant differences were found between the preoperative assessments and the 3 month postoperative follow-up. We repeated the same analyses with z scores, with similar results. In Figure 1 the raw scores as well as the z scores were used to better illustrate whether the mean scores were clinically (≤ -1.5) or pathologically (≤ -2) impaired or not²⁹.

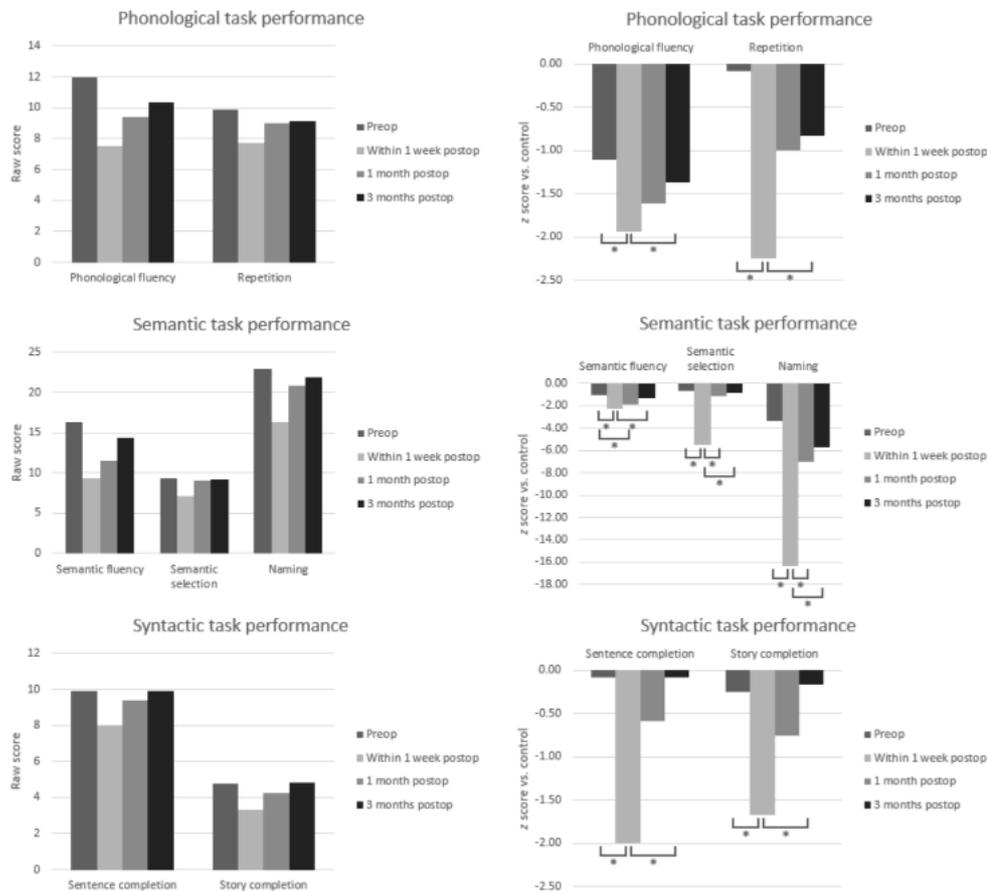


Figure 1: Patients' language performance (in raw scores and z-scores) on the TeleLanguage subtests; * = significant results at an alpha level of < .05

In-person data were missing for three patients and no 3 month follow-up in-person data were available. Consequently, the in-person data were analysed for the remaining 11 patients prior to surgery and at 1 week and 1 month postsurgery. The QAB tests from Table 3 were selected and percentage scores were used to compare the performances at different time points (Figure 2). At 1 week postsurgery, there was a significant decline for all selected QAB tasks except for semantic non-verbal cognition ($p = .48$): QAB naming ($p = .018$), QAB repetition ($p = .011$), QAB sentence production ($p = .028$) and QAB connected speech ($p = .043$). During short-term follow-up (between 1 week and 1 month postsurgery), a significant improvement was observed for QAB naming ($p =$

.018). No significant differences were found between the preoperative assessments and the 1 month postoperative follow-up.

For the TeleLanguage assessments, the non-parametric paired tests were repeated using percentage scores for a selection of the tests described in Table 3 and for 11 patients (for whom in-person data was available). The pattern of results was similar for this set of tasks in this group of patients as it was for the in-person data except for the semantic selection task (Figure 2). There was a significant decline at 1 week postsurgery for verbal naming ($p = .011$), repetition ($p = .041$), semantic selection ($p = .011$), story completion ($p = .027$) and sentence completion ($p = .011$). During short-term follow-up (between 1 week and 1 month postsurgery) a significant improvement was observed for verbal naming ($p = .007$) and semantic selection ($p = .026$). No significant differences were found between the preoperative assessments and the 1 month postoperative follow-up.

In-person data versus telephone data

The in-person data (using percentages) were compared with the TeleLanguage data for 11 patients prior to surgery, at 1 week and at 1 month postsurgery. The Wilcoxon signed-rank test only showed significant differences between telephone data and in-person data for the following task at one test moment: verbal naming versus QAB naming at 1 month postsurgery ($p = .041$) (Figure 2).

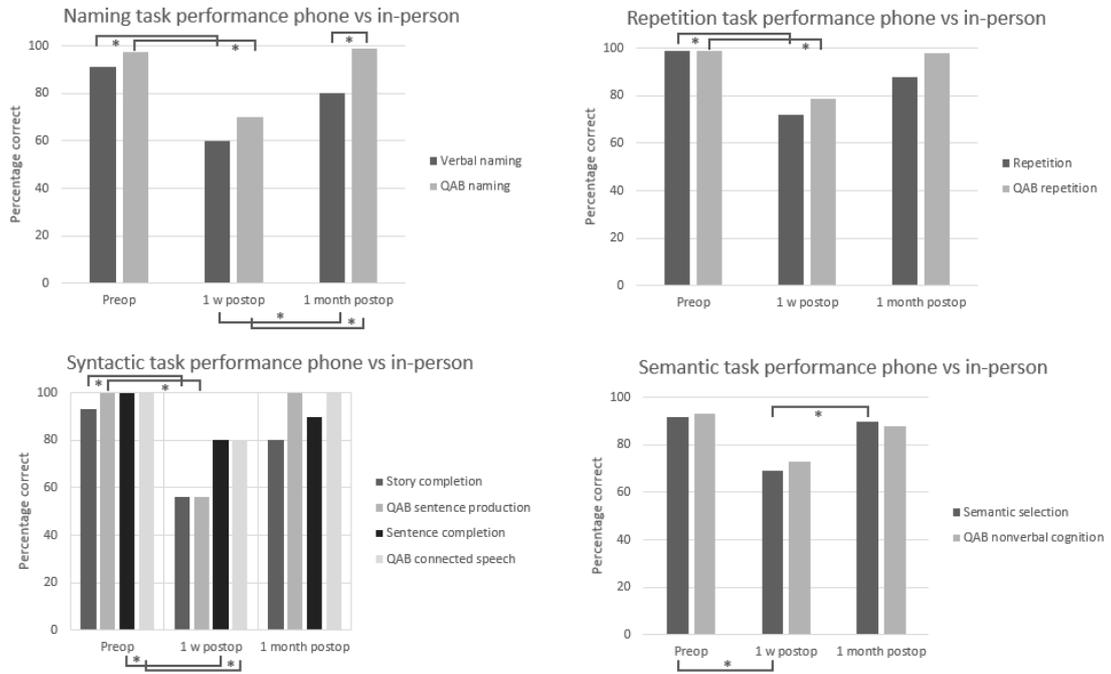


Figure 2: Comparison of in-person language performance and TeleLanguage performance in percentage correct; * = significant results at an alpha level of < .05

DISCUSSION

In this study a telephone-based English language battery (TeleLanguage Test) was developed for the first time in order to follow-up more brain tumor patients undergoing awake surgery. The TeleLanguage Test consists of a naming task, two phonological, two semantic and two syntactic tasks and enables to assess various linguistic functions in a short period of time. In this explorative study, TeleLanguage was administered in a group of 14 brain tumor patients, who also had in-person assessments, and in a group of 14 carefully matched healthy controls. The high feasibility and preliminary validity of TeleLanguage as well as some limitations and future directions will be discussed in the sections below.

Feasibility of telephone-based language battery (TeleLanguage)

Analysis of the data revealed that the TeleLanguage instrument is feasible to assess language functions via the telephone in both healthy controls and brain tumor patients. As no significant differences were found between the two parallel versions A and B, both versions are comparable and so equally reliable to detect language deficits.

In the healthy control group, ceiling effects were found for all tasks except for the fluency tasks. The target answers in the fluency tasks were numerous and there was no maximum score, which induced more variance between subjects than in the other linguistic tasks. The few errors made by the control group in the naming and semantic selection task could not be systematically related to specific items.

In the patient group, the ceiling effects were no longer visible and, as expected, errors were made in multiple tasks (see next section), especially in the postoperative phase.

Clinical application and validity of TeleLanguage

Both pre- and postoperatively, the mean performance of the patients was worse than the performance of the normal population, especially for verbal naming and the fluency tasks as these task scores were still impaired at three months postsurgery. This is in accordance with other studies showing long-term deficits for naming and fluency^{4-6,18,19}.

In the patient group, all TeleLanguage test scores significantly declined shortly (1 week) after surgery with a recovery to preoperative levels at 3 months postsurgery for naming and fluency tasks and a recovery to normal levels for the other language tasks (repetition, semantic selection, story completion, sentence completion). In the study of Wilson et al.¹² the WAB¹³ language tasks revealed language recovery to the presurgical level at 1 month postsurgery except for the naming task. Although a comparable group of brain tumor patients (44% low grade gliomas, 43% LGG in our study) in the same setting (UCSF Medical Center) was included in their study, the following differences might explain the earlier recovery found. First of all, the number of patients in Wilson's study was higher (n = 110) and the group studied included not only brain tumor patients but also patients with epilepsy (n = 13), patients with vascular malformations (n = 5) and other etiologies (n = 4) who might have had higher language scores influencing the total scores. Second, the WAB tests are developed to identify aphasia in stroke patients who typically have more severe language deficits whereas TeleLanguage was composed to detect mild language problems. Therefore, more subtle language deficits might have been overlooked by the WAB tests. Third, other factors such as tumor location, extent of resection and adjuvant therapy might have been

different in both studies and as such these factors might have differentially impacted the results ^{4,5,16,30,31}.

In agreement with many studies, our study shows that naming and semantic fluency seem to have the worst recovery trajectory. Anomia is indeed one of the most pervasive deficits and has been linked to several (posterior) brain regions and pathways ^{32,33}. The multidimensional nature of semantic fluency (lexical retrieval, semantic memory, attention, executive functions) requiring a larger functional network ³⁴ might explain why compensatory mechanisms affect semantic fluency only after restoration of the linguistic and attentional networks. The posterior tumor location in almost two third of our cases (9 out of 14) might also clarify the low scores for semantic tasks and fluency tasks as naming/semantics and attentional and executive functions have been related to temporal and parietal locations. When we compare the four cases with worst language performance (pt 3, 4, 7, 14, Table 2) and the four cases with best language performance (pt 2, 5, 10, 11, Table 2), intraoperative detection of subcortical language sites and concomitant apraxia of speech are linked to the first group and not to the latter. Although the sample size is small, subcortical injuries ³⁵ and concomitant speech problems ³⁶ have shown to negatively influence language scores.

Analysis of the in-person data (until 1 month) revealed a similar profile as described above except for the semantic non-verbal cognition performance, which did not decline after surgery. Only the first 2 items of the non-verbal cognition task were selected as only these items require semantic categorization (as in the TeleLanguage) instead of visual categorization. The small number of items might explain why no significant differences were found between test moments and why the semantic non-verbal cognition task might not be sensitive enough. As the semantic selection task seems to be a sensitive task (significant decline, significant improvement), a longer in-

person assessment version (more items) should be developed in the future. Although the psychometric profile of the naming performance between test moments (pre, post 1 week, post 1month) was similar for both QAB naming and TeleLanguage verbal naming, a significant difference was found between the in-person and telephone data, i.e., significantly higher scores for the QAB naming than for the TeleLanguage verbal naming. As the QAB naming tasks only included 6 items, less low frequent items and no verbs, this naming task might be easier than verbal naming³⁷. Consequently, the difficulty level and not the administration mode (telephone vs in-person) probably explains the significant difference found between both naming tasks.

To conclude, TeleLanguage seems to be a feasible and very practical measure for language assessments in brain tumor patients undergoing awake surgery in a very short period of time. TeleLanguage seems to be a valid alternative for in-person assessments as well. In addition, the verbal naming task and especially the semantic selection task of TeleLanguage are even more sensitive than the in-person measures/alternatives.

Limitations and future directions

First of all, this explorative study was only performed with a small number of patients to pilot the use of TeleLanguage for follow-up of brain tumor patients undergoing awake surgery. No correlational analyses were executed as the in-person tasks used visual stimuli and were therefore very different in nature from the telephone tasks using auditory stimuli. However, as the psychometric profile of the telephone and in-person language performance between test moments was similar, the preliminary validity data are promising and large studies are needed to confirm our findings. In future studies a

control group of patients with a brain tumor in the nondominant hemisphere could be included to take into account generalized surgical effects. Currently, nondominant cases are not referred for pre- and postoperative language assessments.

Second, since most of our patients had mild language deficits, we did not test the application of TeleLanguage in patients with more severe language impairments, which might be more difficult. Until now – in our consecutive series of patients - we have not yet encountered patients with severe language impairments. In the future development of TeleLanguage, the current ‘general comprehension screening’ will be taken as a formal test within the battery to not exclude patients with severe comprehension deficits. For the present manuscript, however, this change would not have impacted our results since in our consecutive series of patients, all patients could complete the comprehension test. However, this battery is especially designed for patients with mild language deficits such as tumor patients³⁸ and might be further clinically tested in other patient populations with mild language problems such as mild head trauma patients, cerebellar patients, etc.

Third, all TeleLanguage tasks had two parallel versions except for the fluency tasks, which might have positively impacted the fluency scores due to practice effects. Nevertheless, as fluency deficits were still found at 3 months postsurgery a learning effect was probably negligible.

Although in our study no significant differences were found for repetition by telephone versus in person - which was the case in Pendlebury et al.²² - detailed analysis of the telephone repetition data is more difficult than for the in-person repetition data. A qualitative error analysis of repetition data into error classifications such as deletion, substitution, etc. is hard if administered via the telephone and no face images are available. Therefore, the repetition task can be part of a telephone-based language test

battery to evaluate language in general, but if the focus of the study is to analyse repetition errors quantitatively and qualitatively, video or in-person assessments are more useful. Although nonword repetition has proven to be sensitive^{39,40}, we have decided to not include it in the telephone battery for the same reason that video or in-person assessments are required to reliably test and score a nonword repetition task. For future data collection, the whole assessment will be recorded, so that we can calculate reaction times, which have shown to add predictive value in addition to correctness⁴¹.

Finally, external factors (i.e., tv, radio, other interruptions) and internal factors (i.e. tiredness, medical illness) and the impact upon a patient's performance during telephone assessment cannot be completely excluded. We attempted to reduce any type of interference by requesting at the start of the session that each participant sat in a quiet room with no other person and turned off all media according to the best practices for remote assessments⁴².

Conclusion

The use of the TeleLanguage battery to conduct language assessments from afar can provide convenience, reduces traveling costs, and optimizes patient care as assessment services might be unavailable locally. In addition, it may be effective in longitudinal clinical research studies to decrease the number of drop-outs as it reduces the burden associated with frequent follow-up in-person testing. The feasibility and preliminary validity of this brief test battery appears to be high. Therefore, TeleLanguage seems to be a valid and reliable tool for various clinical and scientific purposes.

Funding

Elke De Witte is a postdoctoral fellow of the Research Foundation - Flanders (FWO).

VP was funded by the Netherlands Organization for Scientific Research (grant number 446-13-009).

REFERENCES

1. De Witte E, Mariën P. The neurolinguistic approach to awake surgery reviewed. *Clin Neurol Neurosurg.* 2013;115(2):127-145. doi:10.1016/j.clineuro.2012.09.015.
2. Duffau H. Contribution of cortical and subcortical electrostimulation in brain glioma surgery: Methodological and functional considerations. *Neurophysiologie Clinique/ Clinical Neurophysiology.* 2007;37:373-382.
3. De Witt Hamer PC, Robles SG, Zwinderman AH, Duffau H, Berger MS. Impact of intraoperative stimulation brain mapping on glioma surgery outcome: a meta-analysis. *J Clin Oncol.* 2012;30(20):2559-2565. doi:10.1200/JCO.2011.38.4818.
4. Satoer D, Vork J, Visch-Brink E, Smits M, Dirven C, Vincent A. Cognitive functioning early after surgery of gliomas in eloquent areas. *J Neurosurg.* 2012;117(5):831-838. doi:10.3171/2012.7.JNS12263.
5. Satoer D, Visch-Brink E, Smits M, et al. Long-term evaluation of cognition after glioma surgery in eloquent areas. *J Neurooncol.* 2014;116(1):153-160. doi:10.1007/s11060-013-1275-3.
6. Talacchi A, Santini B, Savazzi S, Gerosa M. Cognitive effects of tumour and surgical treatment in glioma patients. *J Neuro-Oncol.* 2011;103(3):541-549.
7. Teixidor P, Gatignol P, Leroy M, Masuet-Aumatell C, Capelle L, Duffau H. Assessment of verbal working memory before and after surgery for low-grade glioma. *J Neuro-Oncol.* 2007;81(3):305-313.
8. Tucha O, Smely C, Preier M, Becker G, Paul G, Lange K. Preoperative and postoperative cognitive functioning in patients with frontal meningiomas. *J Neurosurg.* 2003;98(1):21-31.
9. Wu AS, Witgert ME, Lang FF, et al. Neurocognitive function before and after surgery for insular gliomas Clinical article. *J Neurosurg.* 2011;115(6):1115-1125.
10. Duffau H. Brain plasticity and tumors. *Adv Tech Stand Neurosurg.* 2008;33:3-33.
11. De Witte E, Satoer D, Robert E, et al. The Dutch Linguistic Intraoperative Protocol: A valid linguistic approach to awake brain surgery. *Brain Lang.* 2015;140:35-48. doi:10.1016/j.bandl.2014.10.011.
12. Wilson SM, Lam D, Babiak MC, et al. Transient aphasias after left hemisphere resective surgery. *J Neurosurg.* 2015;123(3):581-593. doi:10.3171/2015.4.JNS141962.
13. Kertesz A. *Western Aphasia Battery Revised. Examiner's Manual.* San Antonio, Texas: Harcourt Assessment, Inc.; 2007.
14. Kaplan EF, Goodglass H, Weintraub S. *The Boston Naming Test, Ed 2.* Philadelphia: Lea and Fibiger; 1983.
15. Wilson SM, Eriksson DK, Schneck SM, Lucanie JM. A quick aphasia battery for efficient, reliable, and multidimensional assessment of language function. *PLoS One.* 2018;13(2). doi:10.1371/journal.pone.0192773.
16. Ilmberger J, Ruge M, Kreth F-W, Briegel J, Reulen H-J, Tonn J-C. Intraoperative mapping of language functions: a longitudinal neurolinguistic analysis. *Journal of Neurosurgery.* 2008;109(4):583-592.
17. Mandonnet E, De Witt Hamer P, Poisson I, et al. Initial Experience Using Awake Surgery for Glioma: Oncological, Functional, and Employment Outcomes in a Consecutive Series of 25 Cases. *Neurosurgery.* January 2015. doi:10.1227/NEU.0000000000000644.
18. Santini B, Talacchi A, Squintani G, Casagrande F, Capasso R, Miceli G. Cognitive outcome after awake surgery for tumors in language areas. *Journal of Neuro-Oncology.* 2012. doi:DOI 10.1007/s11060-012-0817-4.

19. Talacchi A, d'Avella D, Denaro L, et al. Cognitive outcome as part and parcel of clinical outcome in brain tumor surgery. *J Neurooncol.* 2012;108(2):327-332. doi:10.1007/s11060-012-0818-3.
20. Kwan RY-C, Lai CK-Y. Can Smartphones Enhance Telephone-Based Cognitive Assessment (TBCA)? *International Journal of Environmental Research and Public Health.* 2013;10(12):7110-7125. doi:10.3390/ijerph10127110.
21. Mitsis EM, Jacobs D, Luo X, Andrews H, Andrews K, Sano M. Evaluating cognition in an elderly cohort via telephone assessment. *Int J Geriatr Psychiatry.* 2010;25(5):531-539. doi:10.1002/gps.2373.
22. Pendlebury ST, Welch SJV, Cuthbertson FC, Mariz J, Mehta Z, Rothwell PM. Telephone assessment of cognition after transient ischemic attack and stroke: modified telephone interview of cognitive status and telephone Montreal Cognitive Assessment versus face-to-face Montreal Cognitive Assessment and neuropsychological battery. *Stroke.* 2013;44(1):227-229. doi:10.1161/STROKEAHA.112.673384.
23. Unverzagt FW, MONAHAN PO, MOSER LR, et al. The Indiana University Telephone-Based Assessment of Neuropsychological Status: A new method for large scale neuropsychological assessment. *J Int Neuropsychol Soc.* 2007;13(5):799-806. doi:10.1017/S1355617707071020.
24. Wilson RS, Leurgans SE, Foroud TM, et al. Telephone Assessment of Cognitive Function in the Late Onset Alzheimer's Disease Family Study. *Arch Neurol.* 2010;67(7):855-861. doi:10.1001/archneurol.2010.129.
25. Goodglass K, Kaplan E, Barresi B. *Boston Diagnostic Aphasia Examination Third Edition (BDAE-3).* Philadelphia, PA: Lea and Febiger; 2000.
26. Yochim BP, Beaudreau SA, Kaci Fairchild J, et al. Verbal naming test for use with older adults: development and initial validation. *J Int Neuropsychol Soc.* 2015;21(3):239-248. doi:10.1017/S1355617715000120.
27. Goodglass H, Kaplan E. *The Assessment of Aphasia and Related Disorders.* Philadelphia: Lea and Fibiger; 1972.
28. Veale JF. Edinburgh Handedness Inventory - Short Form: a revised version based on confirmatory factor analysis. *Laterality.* 2014;19(2):164-177. doi:10.1080/1357650X.2013.783045.
29. Palmer BW, Boone KB, Lesser IM, Wohl MA. Base Rates of "Impaired" Neuropsychological Test Performance Among Healthy Older Adults. *Archives of Clinical Neuropsychology.* 1998;13(6):503-511. doi:10.1016/S0887-6177(97)00037-1.
30. Ius T, Isola M, Budai R, et al. Low-grade glioma surgery in eloquent areas: volumetric analysis of extent of resection and its impact on overall survival. A single-institution experience in 190 patients: clinical article. *J Neurosurg.* 2012;117(6):1039-1052. doi:10.3171/2012.8.JNS12393.
31. Prabhu RS, Won M, Shaw EG, et al. Effect of the addition of chemotherapy to radiotherapy on cognitive function in patients with low-grade glioma: secondary analysis of RTOG 98-02. *J Clin Oncol.* 2014;32(6):535-541. doi:10.1200/JCO.2013.53.1830.
32. Coello AF, Moritz-Gasser S, Martino J, Martinoni M, Matsuda R, Duffau H. Selection of intraoperative tasks for awake mapping based on relationships between tumor location and functional networks. *J Neurosurg.* 2013;119(6):1380-1394. doi:10.3171/2013.6.JNS122470.
33. Hickok G. The cortical organization of speech processing: feedback control and predictive coding the context of a dual-stream model. *J Commun Disord.* 2012;45(6):393-402. doi:10.1016/j.jcomdis.2012.06.004.
34. Goldstein B, Obrzut JE, John C, Hunter JV, Armstrong CL. The impact of low-grade brain tumors on verbal fluency performance. *J Clin Exp Neuropsychol.* 2004;26(6):750-758. doi:10.1080/13803390490509376.
35. Trinh VT, Fahim DK, Shah K, et al. Subcortical injury is an independent predictor of worsening neurological deficits following awake craniotomy procedures. *Neurosurgery.* 2013;72(2):160-169. doi:10.1227/NEU.0b013e31827b9a11.

36. Papathanasiou D of S& LTTEIWGKPGI, Papathanasiou I, Coppens P, Potagas C. *Aphasia and Related Neurogenic Communication Disorders*. Jones & Bartlett Publishers; 2012.
37. Bastiaanse R, van Zonneveld R. Broca's aphasia, verbs and the mental lexicon. *Brain Lang.* 2004;90(1-3):198-202. doi:10.1016/S0093-934X(03)00432-2.
38. Duffau H. The huge plastic potential of adult brain and the role of connectomics: New insights provided by serial mappings in glioma surgery. *Cortex.* 2014;58C:325-337.
39. Sierpowska J, Gabarrós A, Fernandez-Coello A, et al. Words are not enough: nonword repetition as an indicator of arcuate fasciculus integrity during brain tumor resection. *J Neurosurg.* May 2016:1-11. doi:10.3171/2016.2.JNS151592.
40. Leonard MK, Cai R, Babiak MC, Ren A, Chang EF. The peri-Sylvian cortical network underlying single word repetition revealed by electrocortical stimulation and direct neural recordings. *Brain Lang.* July 2016. doi:10.1016/j.bandl.2016.06.001.
41. Moritz-Gasser S, Herbet G, Maldonado IL, Duffau H. Lexical access speed is significantly correlated with the return to professional activities after awake surgery for low-grade gliomas. *J Neurooncol.* 2012;107(3):633-641. doi:10.1007/s11060-011-0789-9.
42. Luxton DD, Pruitt LD, Osenbach JE. Best practices for remote psychological assessment via telehealth technologies. *Professional Psychology: Research and Practice.* 2014;45(1):27-35. doi:10.1037/a0034547.