Plan Overview

A Data Management Plan created using DMPonline

Title: Adapting Verbal Fluency Tasks for Non-Native English Speakers: A Functional Near-Infrared Spectroscopy Study

Creator:caitlin illingworth

Principal Investigator: Caitlin Illingworth

Data Manager: Li Su, Daniel Blackburn, Caitlin Illingworth

Affiliation: The University of Sheffield

Template: Postgraduate Research DMP (The University of Sheffield)

Project abstract:

It is estimated that currently over 900,000 people in the UK are living with Dementia (Alzheimer's Society, 2024) and, due to the ageing population, this is number is predicted to keep increasing. This increase is predicted to be the greatest among ethnic minorities including South Asian and Black communities (Alzhiemer’s Research UK, 2024). 1 in 5 Brits identify as being part of an ethnic minority (ONS, 2022), yet the cognitive assessments used to diagnose Dementia overwhelmingly based on the norms of White English monolinguals, causing increased rates of misdiagnosis in patients from an ethnic minority and those who speak English as an additional language (Carvalho et al., 2024; Khan & Tadros, 2014; Milani et al., 2018, Ratcliffe et al., 2021).

A common feature of these cognitive assessments are verbal fluency tasks (VFTs) such as naming words that belong to a category e.g. animals (Semantic) or begin with a certain letter (Phonemic). While these tasks are highly sensitive in detecting early signs of cognitive impairment and predicting conversion into Dementia (Cintoli et al., 2024), non-native speakers face additional language barriers on top of the cognitive demands of the test. A non-native English speaker completing a VFT in an NHS memory clinic may score significantly lower on this task than a native English speaker due to:

1) Phonemic barriers: e.g. Arabic native speakers can often incorrectly provide “B-word” responses on a “P-word” task as this letter does not exist in Arabic,

2) Translation barriers: e.g. knowing the names of animals in Mandarin but not their translation into English,

3) A neurodegenerative memory issue.

The ability to disentangle performance due to a language barrier versus neurodegenerative deterioration is critical to accurately identifying patients with memory problems across language backgrounds.

This project will investigate the feasibility of VFTs that are appropriate for non-native English speakers completing the assessment in English, and test similar cognitive processes to when they are completing the assessment in their native lanugage.

VFT performance and brain activity measured through functional near infrared spectroscopy (fNIRS) will be collected from native and Non-native English speakers in their first language and in English. We will use performance and brain activity as a measure to identify and develop more appropriate VFTs for potential memory-clinic patients across language

backgrounds
Through this, we hope to identify more cross linguistically appropriate VTFs to be used as part of memory assessments for dementia to meet the needs of an increasingly multilingual society

ID: 158487

Start date: 01-07-2024

End date: 31-12-2027

Last modified: 14-10-2025

Copyright information:

Adapting Verbal Fluency Tasks for Non-Native English Speakers: A Functional Near-Infrared Spectroscopy Study

Defining your data

- What digital data (and physical data if applicable) will you collect or create during the project?
- How will the data be collected or created, and over what time period?
- What formats will your digital data be in? (E.g. .docx, .txt, .jpeg)
- Approximately how much digital data (in GB, MB, etc) will be generated during the project?
- Are you using pre-existing datasets? Give details if possible, including conditions of use.

This research project will be broken down into 4 work packages over 3.5 years: (1) Piloting verbal fluency tasks and fNIRS for different language backgrounds, (2) VFT+ fNIRS Data collection- Healthy participants, (3) VFT+ fNIRS Data analysis- Healthy participants and (4) Feasibility study with MCI patients.

I predict that work package 1 (4 months) + 2's (20 months) data collection will last around 24 months.

Work package 1: Developing a standardised verbal fluency task and fNIRS procedure

This will be an exploratory piloting of different VFTs and fNIRS normalising techniques with mono and bilinguals from different language backgrounds and ethnicities with the aim of developing a standardised verbal fluency task and fNIRS procedure for work package 2. Previous research indicates that factors such as an individual's hair colour, hair texture, and the melanin of their skin (Kwasa et al., 2023) may effect fNIRS signalling and accuracy. Additionally, as different languages contain different phonemes, and different categories are more relevant to some cultures, verbal fluency tasks can be largely inappropriate (and unusable) with different groups of participants. As such, I will pilot fNIRS and different VFTs with mono and bilingual participants from different ethnicities and language backgrounds. The results of this pilot will inform the VFTs and procedures used during the main data collection.

Consent: Informed consent will be collected before proceeding with the verbal fluency tasks and fNIRS brain recording in the form of electronic consent forms, or pen and paper consent forms as needed. Electronic consent forms received will be stored as a PDF on a University of Sheffield secure Google drive and access will only be given via a password-protected university account to a limited number of researchers directly involved in the study. This is in accordance with the University of Sheffield’s Acceptable Use of Data policy. Pen and paper consent forms will be stored in locked filing cabinets.

Procedure:

Participants will complete a demographics collecting basic demographic information (e.g. age, sex, years of education, ethnicity) as well as specific language background/proficiency data and information on skin tone and hair type.

Audio-only recordings of the Verbal Fluency Tasks completed in the participant's native language (if Bilingual) and in English will be taken with prior consent using an encrypted University Laptop or dictaphone. Audio files will be stored on a password-protected University of Sheffield server. Only audio will be recorded to reduce risk of participant identification.

fNIRS saturation and activity recordings (blood flow in the brain) will be taken simultaneously, with time markers in the recording to match the VFT audio recordings to participants' brain activity. This recording will be collected on a University of Sheffield password protected laptop with the relevant fNIRS recording software.

A short video recording of the placement of the fNIRS cap will be taken for subsequent photogrammetric optode co-registration. This recording will be taken on a digital camera with an SD card, before being uploaded to the X-drive for processing. The recording will be deleted from the SD

card following upload. Once the 3D model is generated through the photogrammetry software, the face (bar cranial landmarks) will be removed from the model, and the recording deleted from the X drive.

Participants will then be asked to complete a brief feedback form to give their opinions of the task/ the setup time/ wearing the fNIRS cap/ and any other aspect of the procedure for implementation in work package 2. e.g. an optional open text single question survey "e.g. Do you have any feedback you would like to give". This will be collected through an online google form, or pen and paper if preferred. This will be stored securely on the University server or in a locked filing cabinet.

All recordings will be uploaded on the University of Sheffield Google Drive as soon as possible, and deleted from the recording device after their quality are double checked.

Interviews will be transcribed and translated by a professional transcriber or by a member of the study team and stored on a password-protected University of Sheffield server once they have signed a confidentiality agreement. Only audio files and a unique participant research number, given following consent, will be shared with the transcribers. Once transcribed, audio recordings will be destroyed as soon as possible to minimise the risk of participant identification. Transcripts will be kept for up to 15 years before destruction.

Sample Size:

Based on the number of participants reported in previous fNIRS piloting, I will aim to recruit around 25 healthy participants, ideally:

- 5 participants with dark brown/black hair
- 5 participants with light brown hair
- 5 participants with blonde hair
- 5 participants with little/no hair
- 5 participants with tight curly hair
- 5 participants with straight hair

Each ppt will complete 20 mins of VFT and rest tasks with concurrent fNIRS brain recordings, so I anticipate the following data requirements:

- Consent/ Demographic forms (.txt)
- Feedback survey (.txt)
- VFT audio recordings (.mp3)
- fNIRS recordings (.nirs)
- VFT transcripts (.txt)

Work package 2: VFT+ fNIRS Data collection- Healthy participants

Following initial piloting in participants with different hair types and ethnicities, a subsequent full-scale experiment will be undertaken with monolinguals and bilinguals from different community centres across Sheffield and Copenhagen.

The methods for Consent collection and storage will be the same as in work package 1. Similarly, the procedure used to obtain VFT and fNIRS data will based on the procedure used in Work Package 1. However, the content of the VFTs and fNIRS set up procedure is informed by Work package 1's findings. Additionally, participants will complete a brief cognitive assessment, such as the MoCA, to verify their diagnostic status.

Sample Size:

Based on the number of participants usually required for a powerful fNIRS analysis, I will aim to recruit the following number of participants:

- 3 community centres each with around 30 participants completing the assessment in their native language and English (Chinese community centre, Shipshape South Asian community centre, Israac)
- 30 monolinguals (English)

As fNIRS requires specialist set up, I will be present for all of the data collection, aided by research champions within the Community centres.

Each ppt will complete ~ 40 mins of VFT and rest tasks with concurrent fNIRS brain recordings, so I anticipate the following data requirements:

- Consent/ Demographic forms (.txt)
- VFT audio recordings (.mp3)
- fNIRS recordings (.nirs)
- VFT transcripts (.txt)

Looking after data during your research

- Where will you store digital data during the project to ensure it is secure and backed up regularly? (E.g. University research storage, or University Google drive)
- How will you name and organise your data files? (An example filename can help to illustrate this)
- If you collect or create physical data, where will you store these securely?
- How will you make data easier to understand and use? (E.g. include file structure and methodology in a README file)
- Will you use extra security precautions for any of your digital or physical data? (E.g. for sensitive and/or personal data)

When conducting interviews away from the university premises (e.g. at community centres) I will ensure additional care is taken with the data while travelling such as storing data through a secure VPN.

The process of recording and storing VFT data, as mentioned in the previous section, will be through a secured device (e.g. encrypted laptop or dictaphone) to then be uploaded on Google Drive immediately and deleted from the recording device after their completeness is confirmed. Any pen and paper recorded information will be kept in a locked filing cabinet for security as necessary.

All the above data collected (VFT audios, fNIRS recordings, VFT transcripts, Photogrammetry video recording, and other data analysis documents) will be stored in the University of Sheffield Google Drive.

Only the lead researcher (Caitlin Illingworth) and primary supervisors (Dr Daniel Blackburn and Prof Li Su) will have access to this data. Select members of the research group will have access to anonymised versions of this data, and GCP trained research champions will aid in the initial collection of data, but not the subsequent analysis.

File structure and methodology details will be placed in a plain text file called README.EXAMPLE FILE NAMES

- VFT_Audio_English_YYYYMMDD_Recruitment_Site_Code e.g. VFT_Audio_English_20240919_SCCC_01
- VFT_Transcript_English_YYYYMMDD_Recruitment_Site_Code e.g. VFT_Transcript_English_20240919_Shipshape_01
- VFT_Audio_Native_YYYYMMDD_Recruitment_Site_Code e.g. VFT_Audio_Native_20240919_SCCC_01
- VFT_Transcript_Native_YYYYMMDD_Recruitment_Site_Code e.g. VFT_Transcript_Native_20240919_Israac_01
- VFT_Demographics_YYYYMMDD_Recruitment_Site_Code e.g. VFT_Demographics_20240919_ShipShape_01
- VFT_fNIRS_YYYYMMDD_Recruitment_Site_Code e.g. VFT_fNIRS_20240919_Monolingual_01

All data with identifiable personal details, including the screening log that shows participants' personal information and corresponding assigned pseudonyms, will be stored in an encrypted folder on the University Google Drive and X Drive, separate from the VFT and fNIRS data, in line with University policies on information security and data protection.

Version control through file naming protocols will enable a traceable data storage record. This will be reviewed regularly to ensure any excess/ unneeded data is removed.

All data storing and security protocols and practices will be reviewed with both supervisors regularly.

In addition, the hardware (i.e. laptop/university computer/dictaphone) used for data collection, processing, and analysis will be encrypted as an extra security precaution.

Storing data after your research

- Which parts of your data will be stored on a long-term basis after the end of the project?
- Where will the data be stored after the project? (E.g. University of Sheffield repository ORDA, or a subject-specific repository)
- How long will the data be stored for? (E.g. standard TUoS retention period of minimum 10 years after the project)
- Who will place the data in a repository or other long-term storage? (E.g. you, or your supervisor)
- If you plan to use long-term data storage other than a repository, who will be responsible for the data?

All VFT recordings will be transcribed by a professional transcriber or member of the research team once they have signed a confidentiality agreement. Once transcribed, audio recordings will be destroyed as soon as possible to minimise the risk of participant identification.

Anonymised versions of the data will be kept for up to 15 years in a repository (University of Sheffield's ORDA) before destruction. This will include materials such as analysed data, details of my methodology and a blank consent forms. Data placed in ORDA would be stored for a minimum of 10 years. The responsibility to destroy this data will be that of the lead researcher of the CognoSpeak project.

The study team may require this data for future publications or to inform/ modify machine learning algorithms following the end of this study.

Sharing data after your research

- How will you make data available outside of the research group after the project? (E.g. openly available through a repository, or on request through your department)
- Will you make all of your data available, or are there reasons you can't do this? (E.g. personal data, commercial or legal restrictions, very large datasets)
- If there are reasons you can't share all of your data, how might you make as much of it available as possible? (E.g. anonymisation, participant consent, sharing analysed data only)
- How will you make your data as widely accessible as possible? (E.g. include a data availability statement in publications, ensure published data has a DOI)
- What licence will you apply to your data to say how it can be reused and shared? (E.g. one of the Creative Commons licences)

Once the project has been completed an anonymised and "Cleaned"- i.e. processed versions of the data will be made openly available with the participant's consent through an online repository for researchers to access.

This will be referenced in a data availiblity statement in future publications.

Putting your plan into practice

- Who is responsible for making sure your data management plan is followed? (E.g. you with the support of your supervisor)
- How often will your data management plan be reviewed and updated? (E.g. yearly and if the project changes)
- Are there any actions you need to take in order to put your data management plan into practice? (E.g. requesting University research storage via your supervisor.)

I will be responsible for ensuring my data management plan is followed as I am the lead researcher of this project. My primary supervisors, Dr Daniel Blackburn and Professor Li Su, will also have access to all the data, and they will support me in putting this plan into practice. Similarly, members of the CognoSpeak team and Community research champions will have restricted access to the data as needed. This DMP will be reviewed every half year.

If there are any significant changes to the project, the appropriate amendments will be made.

To implement this data management plan, I will request access to the University research storage via my supervisor, Dr Daniel Blackburn.