Plan Overview

A Data Management Plan created using DMPonline

Title: Examining the combined effectiveness of language and later literacy interventions on reading outcomes in primary school children.

Creator: Manon Jones

Principal Investigator: Manon Jones

Contributor: Charles Hulme, Cameron Downing

Affiliation: Bangor University

Funder: Economic and Social Research Council (ESRC)

Template: ESRC Template

Project abstract:

Language, reading, and writing skills are core to a child's education. In Wales, a region of the UK that is relatively socioeconomically disadvantaged, dual-language education in Welsh and English presents a significant challenge, and language and literacy outcomes in both languages are consistently rated as poor and declining (Welsh Government, 2016, 2023).

In this project we will first develop resources to improve Welsh primary education at scale via a language intervention programme aimed at 4-5 year olds (reception year) and a literacy intervention programme aimed at 5-6 year olds (Year 1). Second, we will examine whether improving children's reading skills is best achieved by providing the language intervention *followed by* the literacy intervention, compared with receiving only one intervention, building on the premise that language skills are key to literacy development (Bowyer-Crane et al., 2012). Third, practitioner and policy makers' experiences and reflections over the course of the intervention periods will form a comprehensive set of guidelines for bilingual settings, applicable across the UK to children with English as an Additional Language (EAL), given that 20% of children are schooled in a language other than their home language.

The project draws on our successful early language interventions in England (West et al., 2021; Education Endowment Foundation, 2023), and a literacy intervention currently being rolled out in both Welsh and English with older (7-9 year old) children (Downing et al., 2024; Nuffield Foundation, 2022). Both current projects are being supported at scale by the Department for Education in England and the Education Department of the Welsh Government. Our literacy programme is already identified as a key resource for literacy instruction in Wales (Welsh Government, 2023a).

The research program addresses therefore three critical problems with both**local and global** relevance:

1. Evidence-based, comprehensive Welsh language and literacy instruction materials for primary education do not exist.
2. We do not currently know whether poorer readers – in any language – show better literacy outcomes when they receive training in both language *and* literacy skills.
3. Children educated in a language differing from their home language often enter school with weak oral skills, hindering literacy acquisition.

The project uses two successive randomized controlled trials to validate these materials and to compare the effectiveness of combined language and literacy interventions against standalone approaches. Our key measures will capture language and reading outcomes in Welsh. We therefore address the outlined problems as follows:

1. We will develop validated Welsh-language versions of the Nuffield Early Language Intervention (NELI) and an adapted Research on the Instruction of Literacy with Language (RILL) program for younger children.
2. We will examine the particular benefits of combined language and literacy intervention compared to single interventions and a waitlist control group.
3. We will provide guidelines as to the facilitators and barriers to implementing interventions in a bilingual environment, offering insights into best practices for early language and literacy interventions for bilingual children.

The project aligns with the ESRC’s plan to embed place in research activities (UKRI, 2022), addressing socioeconomic and educational disparities in the UK. It also strongly supports government initiatives to improve outcomes for children in **Wales** (the Healthy Child Wales Programme, the Well-being of Future Generations Act (2015), and the Cymraeg 2050 agenda). It aligns with **UK** priorities (Department for Education, 2020) and**global** priorities, notably UNESCO's emphasis on inclusive education that embraces multilingualism.

In this plan, we specify our project partner and service provider OxEd and Assessment (oxedandassessment.com) - from hereonin referred to as 'OxEd'.

ID: 146170

Start date: 01-03-2025

End date: 28-02-2028

Last modified: 04-06-2024

Copyright information:

Examining the combined effectiveness of language and later literacy interventions on reading outcomes in primary school children.

Assessment of existing data

Provide an explanation of the existing data sources that will be used by the research project, with references

The applicants have detailed knowledge of the literature and have carried out extensive searches. There is no extant data that addresses the research questions outlined in this proposal.

Provide an analysis of the gaps identified between the currently available and required data for the research

There is currently no data available examining whether a combined, successive language intervention followed by a literacy intervention improves outcomes to a greater extent compared with either intervention by itself.

Information on new data

Provide information on the data that will be produced or accessed by the research project

We will assess children at five time points on several language and literacy measured obtaining raw and standardised scores. We will also collect demographic data (e.g., gender, age and language skill). These categorical and numeric data will be placed in spreadsheets (.csv format) to facilitate data analysis and anonymised versions will be stored using password protection on research laptops and cloud storage.

We will also conduct semi-structured interviews yielding recording (.mp3 files) that will be uploaded and secured on password-protected laptops and cloud storage. Any personally identifiable information on audio recordings will be removed to ensure anonymity. All audio recordings will be transcribed for further analysis and stored in .txt format.

We will follow the FAIR principles during the whole project to ensure our data is findable, accessible, interoperable, and reusable. Each .csv and / or .txt file will be adequately labelled, and accompanied by readme .txt files containing information on how/when the data was collected will be produced. The readme files will include specific information on the hardware used during data collection.

Quality assurance of data

Describe the procedures for quality assurance that will be carried out on the data collected

at the time of data collection, data entry, digitisation and data checking.

Teachers will be trained to administer the lessons following the lesson plans established in the programme. Trained teachers will conduct all assessments via OxEd software. All training involves rigorous proceedures currently administered by our group and the OxEd team: Teachers receive full training on implementation of the intervention, followed by fidelity checks and regular contact. A similar method is followed for assessments, in which teachers implement language and reading assessments themselves with supervision from researchers if required.

Backup and security of data

Describe the data security and backup procedures you will adopt to ensure the data and metadata are securely stored during the lifetime of the project.

All data will be collected and stored on OxEd servers intitially. OxEd protocols are compliant with GDPR and UK data protection laws. Nevertheless, the data will not contain any identifiable personal information. Data transferred to the research team via data sharing agreement will be stored in password protected laptops and on cloud storage. Each file will be labelled using a systematic naming convention: behavioural data, transcriptions and audio files data will be labelled with the corresponding participant id as well as the session date.

All the data will be automatically backed up in the university's online storage platform. We will also manually check the data on a weekly basis. The backed-up data files will remain unedited for preservation purposes in order to minimise potential loss during pre-processing and analysis. File editing rights of the master copes will only be granted to the data manager. Editable files will be uploaded to Microsoft Teams, accessible only to members of the research project.

Participants who wish to withdraw from the study will have all their data deleted from our storage within a week of their request.

Management and curation of data

Outline your plans for preparing, organising and documenting data.

Each .csv and/or .txt file generated during the project will be adequately labelled (i.e., participant codes, and version number when appropriate). Research assistants will be trained on how to appropriately label the files. A data analyst will update/correct files' names as needed to ensure consistency across files.

Readme. txt files containing information on how/when the data was collected will also be produced. All ReadMe files (in .txt format) will include clear variable/item descriptions and appropriate labels (e.g., definitions of column headings; scoring information; time points) to ensure that the data is clear to understand and easy to be reused by other researchers in the future. Readme files will be labelled similarly to their corresponding data files in order to facilitate cross-referencing.

For the duration of the project, the editable data will be stored in folders on Microsoft Teams, accesible only to members of the project. The folders will be organised according to the data type (i.e., behavioural data, transcriptions, audio files). The uneditable data (master copies), accessible only to the data manager, will follow a similar structure in the university online storage system.

Participants' names will be stored for the duration of the project, allowing their data to be destroyed should they wish to withdraw from the study. All personal data and identification keys will be permanently deleted at the end of the project. Only anonymised versions of the data will be retained.

Difficulties in data sharing and measures to overcome these

Identify any potential obstacles to sharing your data, explain which and the possible measures you can apply to overcome these.

For the purposes of this project, we will have data sharing agreements between participating schools, OxEd and the research team. We do not anticipate any potential obstacles to sharing the .csv and/or .txt data files generated in our project. Informed consent will be a prerequisite for participation. Therefore, we will only share data from participants who provided us with written consent to do so. We will protect their identities by anonymising all the data made publicly available.

Consent, anonymisation and strategies to enable further re-use of data

Make explicit mention of the planned procedures to handle consent for data sharing for data obtained from human participants, and/or how to anonymise data, to make sure that data can be made available and accessible for future scientific research.

We will provide parents and guardians with sufficient information on how the data will be kept confidential and anonymised. Specifically, parents and guardians will be informed that all data generated will be cleaned and processed prior to any statistical analysis to ensure that no identifiable information is included in any reports, academic articles, or publicly available data repositories. During the cleaning and process of the raw data, we will remove any information on (1) participants' name (and use randomly generated ID numbers instead), (2) participants' school name (and assign a nondescript label to each school instead), and (3) care will be taken to store participant codes.

Parents and guardians will be informed that anonymised data files will be made publicly available at the UK Data Service's online repository (http://reshare.ukdataservice.ac.uk) and in a GitHub repository (https://github.com/) at the end of the project. Parents and guardians will be given the option of requesting, at any point during the project, that their children's data not be made publicly available without having to give a reason.

Our uploading of the data files and corresponding metadata will follow the FAIR data principles to ensure our data is findable, accessible, interoperable and reusable.

Copyright and intellectual property ownership

State who will own the copyright and IPR of any new data that you will generate.

All of the processed data belong to the principle investigator, the co investigators, and the universities

with which they are associated. The data generated in the project will be published under a Creative Commons Attributions Licence (CC BY).

Responsibilities

Outline responsibilities for data management within research teams at all partner institutions

The PI Jones and CI Downing will have primary responsibiltiy over data management. These responsibilities will include quality assurance: managing data collection, extracting data and appropriate data filtering and analysis, appropriate formatting and concatenating of data files, validity checks, secure and well-organised storage.

Data management will be overseen by PI and CIs and the end of the project, these individuals will archive the data: UK Data Service's online repository (https://reshare.ukdataservice.ac.uk/) following the FAIR data principles, as well as in a GitHub repository (http://github.com/).

Preparation of data for sharing and archiving

Are the plans for preparing and documenting data for sharing and archiving with the UK Data Service appropriate?

We wil deposit all of the processed data along with corresponding ReadMe files at the Data Service's online repository within three months of the project ending. All the processed files will be labelled meaningfully and organised in zip bundles being uploaded to the repository.

All ReadMe ffiles (in .txt format) accompanying each spreadsheet (.csv format) will include clear variable/item descriptions and appropriate labels (e.g., definitions of column headings; scoring information; time points) to ensure that the data is clear to understand and easy to be reused by other researchers in the future.

We will process and analyse all the quantitative data in STATA and RStudio. All scripts generated for statistical analyses and data plots will be made available in GiHub, along with all of the relevant .csv files. All of the code included in the scripts will be annotated accordingly to facilitate understanding, and to ensure transparency and reproducibility, in line with Open Science practices.

We will ensure all academic publications based on data resulting from this grant will include information on how to access the supporting data files and their associated metadata, following the FAIR principles.

Is there evidence that data will be well documented during research to provide highquality contextual information and/or structured metadata for secondary users?

Data from our original RILL project, funded by the ESRC, and with similar structural properties to the current proposal, is available here: https://reshare.ukdataservice.ac.uk/855333/