DEVELOPING A CLINICAL LANGUAGE ANALYSIS TOOL TO IDENTIFY LANGUAGE DISORDER:

Capturing samples of children's storytelling through citizen science and mobile technology.

Rebecca Bright[1], Swapnil Gadgil[1], Prof Cristina McKean[2], Dr Yvonne Wren[3], Elaine Ashton[2], Dr Sam Harding[3].

1. Therapy Box 2. Newcastle University 3. Bristol Speech and Language Therapy Research Unit

Background

Speech and language therapists (SLTs) use assessments of children's storytelling to support the diagnosis and treatment of language disorder. However, this is labour intensive involving manual transcription of the child's spoken output analysis of language complexity and comparison to normative data. An NIHR i4i funded collaboration between industry, clinical and academic partners is designing, developing and evaluating an application for SLTs: 'Language Explorer'. This app will record a child's storytelling in a controlled context, uses speech recognition to aid rapid transcription and complete an analysis of the sounds, words, sentences, grammar and story structures the child uses.

Aim

To gather normative data and provide data for the machine-learning driving the language analysis functions a novel citizen science project was undertaken.

Methods

A mobile app was co-designed with clinicians and children within the target age range.
The app was designed, deployed and tested for a range of mobile devices and tablets using the PlayStore and AppStore. It was enabled so that it could be used offline, where the sample would only be uploaded when the user went online.

A proportionate stratified sampling technique was used and stratified with reference to age bands (six bands from 4;0-7;11), gender and five socio-economic status categories.
This was supported by targeted marketing and social media activity.

Manual transcription and analysis of 600 language samples was undertaken and compared with automated analysis and the semi-automated transcription process for agreement level and time taken.

Results

In total 4435 participants created a profile on the citizen science version of the app, and of that 2662 provided a recording following the presentation of the story. The stratified sampling goals were achieved across each sub-group within the sampling framework; having been graded to meet audio quality levels sufficient for transcription and analysis.

There were a significantly higher number of recordings received from SES1 (lowest SES quintile representing the most deprived postcodes) than SES5 (the highest SES quintile). This suggests that using mobile technology increased the ability for low SES households to participate in research based citizen science approaches for data collection. Further exploration of the geography of participants is planned.

Demographics of participants who provided an audio narrative sample in the app

Citizen science participants

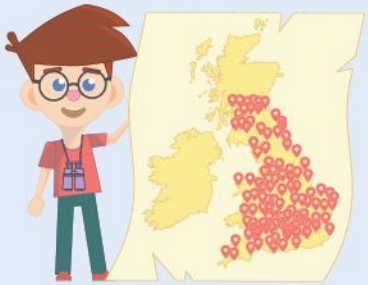
Country	Profiles made in the app	Audio sample recorded	Meeting eligibility criteria for stratification
Great Britain	4138	2509	1481
Rest of the world	297	153	0

Eligibility criteria - aged between 4;0 and 7;11, English spoken as primary language, no reported disabilities.

Socio-economic status of the British participants who did record samples

Socioeconomic status	Audio samples received
SES 1	561
SES 2	527
SES 3	477
SES 4	524
SES 5	400
other	20

Socio-economic status was determined using postcode-assigned Index of Multiple Deprivation score represented in quintiles.



Map of participants in the UK

Discussion

Using a citizen science approach to data collection for the purposes of informing the development of a future clinical tool for children is a time effective and feasible process. Quality of data collected via mainstream mobile phones is sufficient for the purposes of language sampling. Importantly collection of more than six times the target number of samples were required to fulfill our stratification strategy.

This study/project is funded by the National Institute for Health Research (NIHR) Invention for Innovation programme. The views expressed are those of the author(s) and not necessarily those of the NIHR or the Department of Health and Social Care.
Citizen science work package approved by University of Bristol Faculty of Health Science Research Ethics Committee Ref: 97304

Web address - https://www.languageexplorer.app/study | Email address - rbright@therapy-box.co.uk