The Reading Brain Project: An Open Science Data-Sharing Initiative

Anya Yu[1], Benjamin Schloss[1], Chun-Ting Hsu[2], Lindsey Ma[1], Chih-Ting Chang[3], Marissa Scotto[1], Rike Seyfried[1], and Ping Li[1]

[1]Department of Psychology, The Pennsylvania State University, USA, [2]Kokoro Research Center, Kyoto University, Japan, [3]Institute of Linguistics, Academia Sinica, Taiwan

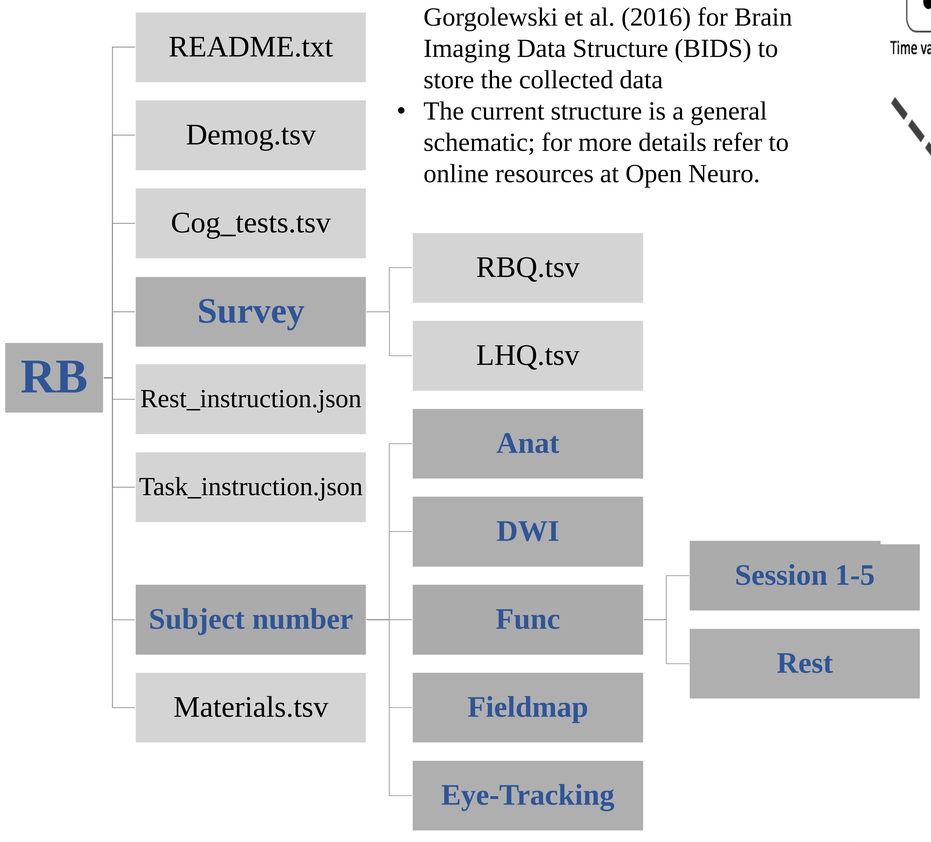
Background & Aim

- There's been a surge of neuroimaging studies on reading; however most are limited to small sample sizes and lack a common task.
- We aim to understand the neurocognitive mechanism underlying text comprehension by adult native (L1) readers, adult non-native second language (L2) readers, and native middle-school readers.
- We provide a framework for how to use this dataset to test theories in text reading, and to compare against datasets that are smaller in size

Participants

	L1 adults	L1 children	L2 adult	
Location	Penn State Hershey	State College	State College	Beijing
N	50 (26F)	54 (28F)	28 (9F)	28 (21F)
Age	22.88 (4.71)	12.8 (.75)	24.36 (3.90)	25.93 (5.41)

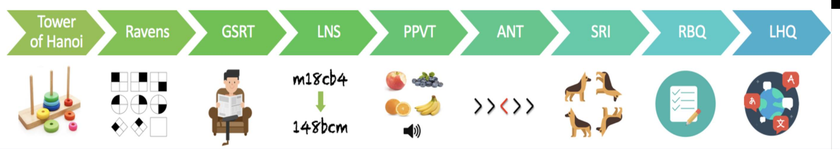
Data Structure



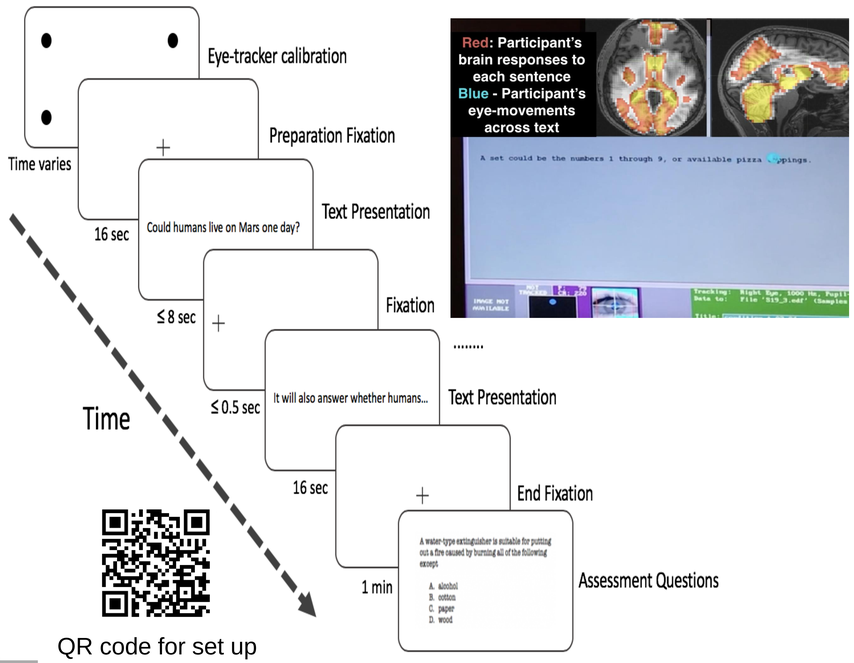
- We followed the principles of Gorgolewski et al. (2016) for Brain Imaging Data Structure (BIDS) to store the collected data
- The current structure is a general schematic; for more details refer to online resources at Open Neuro.

Procedural Overview

- Session I: Image acquisition
  - Structural data, resting-state fMRI, fixation-related fMRI, DTI
- Session II: Behavioral testing
  - Executive function (Tower of Hanoi, LNS, ANT)
  - Reading (GSRT, PPVT)
  - Individual differences (Raven's, SRI, RBQ and LHQ)

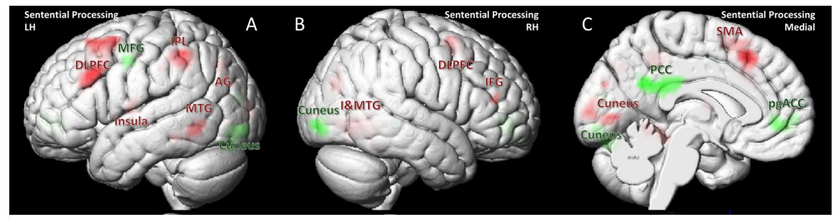


Acquisition



- Structural and resting-state fMRI data: collected to convention
- In scanner, participants were shown 5 expository STEM text sentence by sentence in a self-paced manner.
- Acquisition time varied on the speed of self-paced reading with a max of 5 minutes per reading. All FiRe fMRI data was collected at a multiband factor of 6 (400 ms for each volume acquisition), and spin echo sequence images were collected to calculate distortion correction for multiband sequences.

Data Utilization



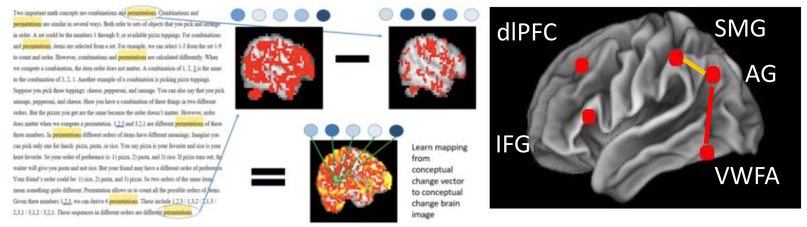


Figure 3. Calculation of conceptual change image.

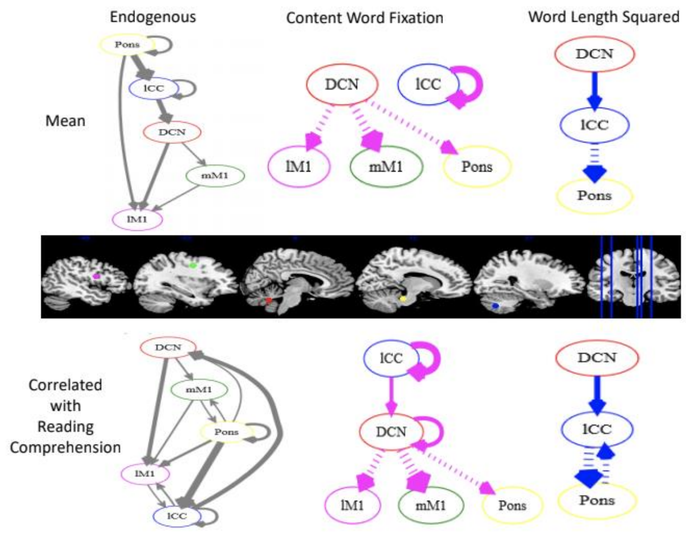


Figure 1. Displays the connectivity results based on the average connectivity between regions and connections that are correlated with reading comprehension scores. Good comprehenders have more interconnected networks both between cortex and cerebellum and brainstem, between the cerebellum and the brainstem, and within the cerebellum itself.

Conclusions

- This project is a distinct dataset with unique methods, encompassing experienced L1 readers, experienced L2 readers (in immersive and non-immersive environments), as well as inexperienced L1 readers of STEM text.
- Potential usage include providing baseline comparisons & providing more statistical power for analysis using machine learning for classification. Providing valuable source of data for gaining insights into language acquisition, reading comprehension, as well as knowledge acquisition.
- We stand by the vision that data sharing is a normal and essential part of the scientific process, and encourage usage of the Reading Brain Project dataset.

References & Resources

- Data and detailed methodology procedures are available on OpenNeuro, please refer to link in http://blclab.org/reading_brain. Detailed references can be found in the document as well

Contact: yuu.anya@gmail.com

Acknowledgments

We would like to thank Jennifer Legault, Friederike Seyfried, and Jing Yang for assisting with the current research project. This project is generously supported by the National Science Foundation (#BCS-1533625) as part of the Integrative Strategies for Understanding Neural and Cognitive Systems (NCS) program., and part of the *PIRE: Bilingualism , mind, and brain* grant.