# The role of text exposure on language comprehension

Public registration  Updates

Metadata

**This is an update to the original registration**

This update was made on Oct 30, 2022

Reason for update:
*Cut off trimming criteria were adjusted based on prior literature - originals were based on single word self-paced reading the updated cut offs reflect whole sentence reading -the type of presentation used in this project.*

## Study Information

**Hypotheses**

2 main effects and an interaction. First, experienced readers (individuals with higher levels of text exposure) should be faster than less experienced readers. Second, sentences that are established in the literature to be harder (take longer) to comprehend should also take longer to comprehend in the present study (object relative clause sentences > subject relative clause sentences > passive main clause sentences > simple active sentences). We hypothesise an interaction such that the individual differences in reading times should be more pronounced in more complex sentences versus simple sentences. More specifically, participants with higher ART and Vocabulary scores should be faster and more accurate with the harder to comprehend sentences which are less frequent in conversational speech and are found mainly in the written language. Participants with higher ART and Vocabulary scores may show less of an advantage for the easier to comprehend sentences, where more adult readers may show ceiling performance.

Additionally we explore whether attitudes towards reading as captured by the survey of reading enjoyment predict the speed and accuracy across the sentence types. The survey of reading enjoyment was previously developed and used to assess reading attitudes in children (Alderson et al., 2019). We are attempting to adapt and use this task for the first time with adult populations. We attempt to investigate whether this survey may be a useful tool to assess reading experience in adult populations.

## Design Plan

**Study type**

ıts. This is also known as an intervention experiment



**Is there any additional blinding in this study?**

*No response*

**Study design**

Latin-square design for the self-paced reading study. Total 120 sentences were split into 2 lists. 20 simple sentences and 20 passive sentences were the same in both lists. A total of 80 subject and object relative sentences with response patterns validated in earlier studies (Traxler, 2002) were used to form 40 sentence pairs as shown in 1 a-d
1a. The banker that the lawyer irritated played tennis every Saturday and Sunday
1b. The banker that irritated the lawyer played tennis every Saturday and Sunday
1c. The lawyer that the banker irritated retrieved the paperwork from the office.
1d. The lawyer that irritated the banker retrieved the paperwork from the office.
These sentence quadruplets were split across 2 lists in the following counterbalancing fashion: List 1: ad; List 2: bc and pseudorandomized within the list with at least 4 items between a and d
Author Recognition Test (ART)(Acheson, Wells and MacDonald, 2008; Moore & Gordon, 2015;Stanovich & West, 1989);
Vocabulary Test (Shipley, 1940)
Self-Perception of Reading Enjoyment (adapted from Alderson et al., 2019)

- Item List_withCode.xlsx
- ART Key.xlsx
- Shipley_Answers.xlsx
- SelfPerceptionReadingEnjoyment_Adults.docx

**Randomization**

Participants are assigned to one of the two lists based on the last number of their birth day. List 1 if the last number of their birth day is even and List 2 if the last number of their birth day is odd. E.g. if one's birthday is on April 14th the person will be assigned to List 1 if the birthday is on April 13th then to List 2.

# Sampling Plan

**Existing Data**

Registration prior to analysis of the data

**Explanation of existing data**

As of the date of submission of this research plan for preregistration, the pilot data has been collected and analysed to test the experimental software and the data collection flow. Several questions were edited due to misspellings. The data collection began on December 10th, 2020 and is ongoing but the data will not be viewed or analyzed until the required number of participants is reached as stated in Sample size description and this pre-registration document.

**Data collection procedures**

Participants are recruited through the University subject pool with the help of SONA software. Participants are granted 1 credit hour for agreeing to participate. Participants must be at least 18 years old.

We will only include participants who learned English before age 5. However, all participants in the research pool will be able to participate, and we may analyze the data of participants who learned English after age 5 as exploratory analyses.

*No files selected*

**Sample size**

Our target sample size is 120 participants. We will attempt to recruit up to 150, assuming that not all will complete the total task, or will have learned English after age 5.

**Sample size rationale**

We used the software simr package in R (green & MacLeod, 2016) to conduct a power analysis. Our goal was to obtain .95 power to detect a small to medium effect size of up to .25 at the standard .05 alpha error probability. Sample size of the pilot = 80. For the power analyses we ran a simulation on the pilot data with the same LME models that we plan to use for the real analyses as described in Confirmatory analyses (section 18 of the "Analyses plan")

**Stopping rule**

We will stop when we get to our N*1.25 because we expect to eliminate participants who learned English after age 5, plus any non-compliance/technical issues.

# Variables

**Manipulated variables**



ipulated across items as described in 7.

Level 4 - 40 (20 seen by each participant) object relative clause sentences;

*No files selected*

**Measured variables**

Whole sentence reading time in milliseconds as recorded by IBEX farm software
Question accuracy - 0 for wrong answer 1 for correct
Text exposure based on the number of authors marked correctly in the ART test 1 point for every author recognized correctly -1 for the wrong selection (i.e. name that does not correspond to an actual author)
Vocabulary - 1 point for every correctly chosen synonym for the target word.
Reading enjoyment survey - a composite score based on the answers to the 10 questions 7 point Likert scale each

*No files selected*

**Indices**

An exploratory analyses following Moor & Gordon (2015) and our judgement to distinguish between literary and popular books to examine whether/how much such division explains the variance within the ART score

*No files selected*

# Analysis Plan

**Statistical models**

LME maximal fit fashion starting with most complex random intercept and slope structure and reducing the complexity of the random structure if the most complex one does not converge

Confirmatory analyses:
Reading time = Sentence type *ART
Accuracy = Sentence type *ART

*No files selected*

**Transformations**

Dummy coding for sentence types to insure orthogonality of the comparisons:
First compare the two easier sentences to the two harder sentences, then compare within each easy/hard group:
Simple Passive SRC ORC
-1 -1 1 1
0 0 -1 1
-1 1 0 0

Linear trend dummy coding scheme:
Simple Passive SRC ORC
-3 -1 1 3

ART, Vocabulary and Reading Enjoyment will be centered and z-scored by participant

**Inference criteria**

We will use the standard p<.05 criteria for determining if the results are significantly different from those expected if the null hypothesis were correct. The Bonferroni adjustment will be used to adjust the p value for multiple comparisons.

**Data exclusion** **Updated**

Two criteria:
1 Exclude people who learned English after 5 years of age and/or had a history of reading difficulties.
2 For the response times a two-step process will be used:
Reading times faster than 1500ms and slower then 300000ms will be excluded additionally the individual reading times will be trimmed with the cut-off of 2.5 standard deviations above and below the individual conditional mean.

**Missing data**

Missing data will not be analysed

**Exploratory analysis**

1. Correlations between Reading enjoyment survey results and the rest of the measures is an exploratory analysis.
2. We are excluding the participants who learned English after the age of 5. Analyses of such participants will not be included in the main analyses but we plan to use exploratory analyses on these subsets to further explore how the reading measures predict rts in the excluded population.