

General Study Information

Title

The open science of reading: A systematic review

Description

The present study is a meta-scientific systematic review of published research literature in the content area of reading in order to estimate the prevalence of participation in key open science practices amongst authors publishing on this topic. This study is a conceptual replication and extension of previous studies by Cook et al. (2023) and Hardwicke et al. (2020; 2021). Rather than estimating the prevalence of open science practices among scientific literature authored by researchers in a particular discipline (e.g., special education or psychology), our review focuses on literature published on a particular topic in 2023: the scientific study of reading, anchored in the Simple View of Reading (Hoover & Gough, 1980).

Study Type

Observational Study / Systematic Review

Study Design

The present study is a cross-sectional systematic review of scientific literature published in 2023.

Research Questions & Hypotheses

Primary Research Question:

1. What is the estimated prevalence of key open science practices in published literature in the subject area of reading in 2023?

Hypotheses

None - This research question is descriptive in nature and claims made will be observational.

Exploratory Research Question:

2. How does the prevalence of open science practice vary by journal, author, and study-level predictors?*

Hypotheses

H0: Prevalence of each open science practice will be equal by journal-, author-, and study-level predictors

H1: Prevalence of each open science practice will vary by journal-, author-, and study-level predictors

* This research question is intentionally broad as it is unclear whether sufficient data will be obtained to conduct the planned analyses for each individual open science practice and/or predictor.

Data Collection

Existing Data

At the time of this pre-registration, a search has been conducted and titles and abstracts of potentially eligible articles have been screened for eligibility; however, the final sample of articles has not yet been randomly selected, nor has any member of the project viewed the full text articles or examined articles for the presence of open science practices or predictors.

Data Collection Procedures

Data collection and coding will occur in three phases.

Phase I: Database Search

In February 2024, a search was conducted in the following two databases: APA PsycINFO and ERIC. For each database, the following search command was used: "decoding" OR "reading comprehension" OR "listening comprehension" OR "language comprehension". The search was limited to articles published in the year 2023 and to include only peer-reviewed articles. These search terms were chosen to return articles aligned with any one or more elements of the Simple View of Reading (Hoover & Gough, 1990). The search returned 1201 total records across both databases after removing 175 duplicates.

Phase 2: Title and Abstract Screening

In February 2024, articles were reviewed with respect to the following exclusion criteria. Articles were removed from the population of eligible articles if they appeared to meet any of the exclusion criteria described in Table 1. Following title and abstract screening, 705 eligible articles remained.

Table 1 Exclusion criteria.

Exclusion criteria	Justification	Records Removed
Article does not have a methods section	Article unlikely to be empirical research study	70
Study does not involve humans (e.g., computer models, animals)	Article does not fit the aims of the study	28
Study clearly not related to any aspect of the Simple View of Reading	Article does not fit the aims of the study	75
Study focuses on a medical patient population or context	Authors of research in medicine may face unique barriers to engaging in open science practices (e.g., privacy concerns, data formats)	35

Study involves neuroscience methodology (e.g., fMRI, EEG, ERP)	Authors of research using these methodologies may face unique barriers to engaging in open science practices (e.g., unique data formats)	288
Total		496

Phase 3: Coding

A team of graduate students will be trained to code articles. Coding materials have been adapted from those shared by Cook et al. (2023). Coding will occur via two Qualtrics surveys: one for coding journal-level predictors and one for coding article and author-level predictors. Version 1 of each coding form is included as part of this pre-registration. However, due to the iterative nature of the coding procedure, minor changes may be made based on coder feedback after the test batch described below. Any changes will be described in the final manuscript.

Three hundred articles will be randomly selected for coding. Twenty additional articles will be selected as back-ups in case any of the initial 300 articles are found to be ineligible during the coding process. Additionally, a list of unique journals associated with the 300 articles will be exported for coding.

Coding will proceed as follows, guided by Stock (1994)'s guidelines. First, the entire coding team will meet with the lead author. Each item on each coding form will be reviewed and discussed, then the team will code several articles and journals as a group. Each coder will then be given one week to independently code 3 articles (or journals). If coders achieve 90% or greater agreement with the lead author, they will be approved to code for the study. If agreement is less than 90%, additional training and discussion will be provided and the coder will independently code 3 new articles (repeated as necessary until 90% agreement is achieved). Coding will then proceed in three batches to guard against coder drift. After each batch of articles, the entire coding team will meet to discuss and resolve any issues as a team. Coding will be complete after the third batch when all 300 articles and their associated journals have been coded and double-coded as assigned.

Sample Size and Rationale

Three hundred articles will be randomly selected for coding. It is currently unknown what the final number of unique journals will be. This sample size is similar to that of Cook et al. (n = 250).

Randomization

300 articles will be randomly selected for coding from the population of articles obtained via the search procedures described below. 150 (50%) of these will be randomly selected for double coding. Random selection will be performed in R using the randomizR package (Coppock, 2023).

Manipulated Variables

This study is observational in nature and no variables will be manipulated.

Measured Variables

Outcomes:

1. Presence of 9 open science practices: open access, open data, open materials, open code, pre-registration, registered report, conflict of interest statement, funding statement, replication*

* Note that “open peer review”, which was coded for the Cook et al. (2023) study, was not coded for the current study due to 0% prevalence observed in the previous study.

2. If open data/materials/code are shared, additional details will be collected about the specifics of how they are shared.

Predictors:

1. Author-level [corresponding or first author if no corresponding author identified]

- 1a. Total number of authors
- 1b. Location
- 1c. Institution type

2. Journal-level:

- 2a. Presence of open access publishing option
- 2b. Presence of open science incentive
- 2c. 2022 journal impact factor
- 2d. Year journal founded

3. Study-level:

- 3a. Study type
- 3b. Sample size
- 3c. Participant populations represented
- 3d. Participant age groups represented
- 3e. Content area (key words)

Analysis

Data Analysis

Reliability

Cohen’s Kappa will be calculated for all double-coded items for each overall coding form and for each section of each coding form.

Descriptive Analysis (RQ1)

As in the prior studies we are conceptually replicating, data will be analyzed descriptively (e.g., calculating and visualizing the prevalence of each open science practice).

Exploratory Analysis (RQ2)

Chi-square testing and Poisson regression will be used to compare the prevalence of open science practice by each predictor for which there is sufficient data. For each, p-values will be reported and compared against a criterion of .05 to determine significance. As mentioned above, the analysis plan for the exploratory analysis is vague as it is unclear whether sufficient data will exist to conduct planned comparisons for each predictor and outcome.

Data Exclusion

Articles will be excluded if they meet any of the exclusion criteria described above. These articles will be replaced from the sample of 20 randomly selected replacement articles as needed.

Acknowledgements and References

As stated previously, this study is a conceptual replication of Cook et al. (2023), which itself is a conceptual replication of Hardwicke et al. (2020; 2022). Most of the materials used in the present study were adapted from those shared by Cook et al., including the coding forms, R scripts, and analysis plan.

References

- Cook, B. G., van Dijk, W., Vargas, I., Aigotti, S. M., Fleming, J. I., McDonald, S. D., ... & Johnson, R. M. (2023). A Targeted Review of Open Practices in Special Education Publications. *Exceptional Children*, 89(3), 238-255.
- Coppock, A. (2023). randomizr: Easy-to-Use Tools for Common Forms of Random Assignment and Sampling. R package version 1.0.0. <https://CRAN.R-project.org/package=randomizr>
- Hardwicke, T. E., Thibault, R. T., Kosie, J. E., Wallach, J. D., Kidwell, M. C., & Ioannidis, J. P. (2022). Estimating the prevalence of transparency and reproducibility-related research practices in psychology (2014–2017). *Perspectives on Psychological Science*, 17(1), 239–251. <https://doi.org/10.1177/1745691620979806>
- Hardwicke, T. E., Wallach, J. D., Kidwell, M. C., Bendixen, T., Crüwell, S., & Ioannidis, J. P. (2020). An empirical assessment of transparency and reproducibility-related research practices in the social sciences (2014–2017). *Royal Society Open Science*, 7(2), 190806. <https://doi.org/10.1098/rsos.190806>
- Hoover, W. A., & Gough, P. B. (1990). The simple view of reading. *Reading and Writing*, 2, 127-160.

Pre-Registration v1 (26 February 2024)

Stock, W. A. (1994). Systematic coding for research synthesis. *The Handbook of Research Synthesis*, 236, 125-138.