

The title

First Author¹ & Ernst-August Doelle^{1,2}

¹ Wilhelm-Wundt-University

² Konstanz Business School

Modul 6b: Empirisch-Experimentelles Praktikum

Dr. Sarina Schäfer

07.08.2023

Author Note

Add complete departmental affiliations for each author here. Each new line herein must be indented, like this line.

Enter author note here.

The authors made the following contributions. First Author: Conceptualization, Writing - Original Draft Preparation, Writing - Review & Editing; Ernst-August Doelle: Writing - Review & Editing, Supervision.

Correspondence concerning this article should be addressed to First Author, Postal address. E-mail: my@email.com

Abstract

One or two sentences providing a **basic introduction** to the field, comprehensible to a scientist in any discipline. Two to three sentences of **more detailed background**, comprehensible to scientists in related disciplines. One sentence clearly stating the **general problem** being addressed by this particular study. One sentence summarizing the main result (with the words “**here we show**” or their equivalent). Two or three sentences explaining what the **main result** reveals in direct comparison to what was thought to be the case previously, or how the main result adds to previous knowledge. One or two sentences to put the results into a more **general context**. Two or three sentences to provide a **broader perspective**, readily comprehensible to a scientist in any discipline.

Keywords: keywords

Word count: X

The title

Methods

Preregistration and version control

The hypotheses, the inclusion/exclusion criteria, used databases, search queries and the basic theoretical foundation of this systematic literature review are preregistered and can be found on Moodle or in the GitHub repository.

As suggested by Lakens (2022) (Chapter 14), the present systemic literature used a GitHub repository to store all data and files. The repository is available at:

https://github.com/julianrottenberg/Stereotype_Threat_im_akademischen_Kontext

This approach allows for more transparency and reproducibility as well as accountability.

Artificial Intelligence (AI)

It should be acknowledged that artificial intelligence has been used as an aid in this review - namely, Anthropic's Claude AI 3.5 Sonnet (Anthropic, 2024) and GitHub's Copilot (GitHub & OpenAi, 2024), the latter was directly integrated into RStudio Server (Posit team, 2024). The chats that directly influenced this review are all available on the GitHub repository. For Github Copilot the autocomplete-style suggestions were used.

Claude AI 3.5 Sonnet was used to generate descriptions of the papers used in this review - based on a template. The process here was as follows: First the template was manually filled out by a human, after this process was completed for every paper, a second template was created, the contents of which were filled out by AI and then, later, used in conjunction with the manually created templates. When the different templates differed from one another the primary source (i.e. the paper the template was based on) was checked again. Both, the human-generated as well as the AI-generated templates can be found on the GitHub repository - the AI generated summaries have been marked as such, beginning with "Claude_Ai_" in their file name.

To clarify, AI was not used to generate any of the text in this review, it was used as a

tool to gather a better understanding and overview of the papers involved. The process of having a human and AI created summary of each paper was chosen to gather an extra layer of security regarding the contents of each paper as well as to counteract possible oversights.

Databases, search queries and inclusion/exclusion criteria

The databases used were Web of Science, Google Scholar, PSYINDEX, ResearchRabbit and EBSCOhost. Within EBSCOhost, the databases APA PsycArticles, APA PsycInfo, Psychology and Behavioral Sciences Collection, PSYINDEX Literature with PSYINDEX Tests, Education Source Ultimate, and Academic Search Ultimate were searched.

Furthermore, the snowball method was utilised to find additional papers - however, this approach did not deliver any additional papers, the same applies to ResearchRabbit.

The permalinks to each search used can also be found within the GitHub repository.

Within Web of Science the included document types were “Article”, “Other”, or “Clinical Trial”; the excluded document types were “Book”, “Meeting”, “Editorial Material”, or “Review Article”. Furthermore, the database “Preprint Citation Index” was excluded.

In EBSCOhost, “Apply equivalent subjects” was applied as an Expander, while “Peer Reviewed”, “Document Type*”, and “Publication Type*” were used as Limiters.

In Google Scholar, the following was added at the end of the search query: “AND”empirical study” AND “peer-reviewed” -books -meta-analysis)“.

These extra filters were applied in accordance with the inclusion and exclusion criteria outlined in the preregistration. No other changes were made to the search queries. An overview of the search queries can be found in Table 1.

The inclusion and exclusion criteria specified in the preregistration were applied to each paper. The criteria “Stereotype Threat”, which required studies to “explicitly examine, manipulate, or measure stereotype threat as a key study variable or factor” was enforced on a lot of papers and resulted in their exclusion - even when they were otherwise relevant (more on this in the discussion section), same applies to the “Outcomes” criteria, which

required studies to report “at least one of the following: 1. Neural activation patterns/brain imaging data; 2. Cognitive processes (e.g., working memory, cognitive control/executive functions)” also resulted in the exclusion of papers which indirectly measured these outcomes but/or did not specifically focus on “working memory” for example - as an example: a paper might have used a test that is known to measure working memory but did not mention “working memory” within its abstract, methods or results section, so it was excluded.

Screening

The screening process was done using the software Rayyan (Ouzzani et al., 2016). All results were imported onto the platform. The total amount of papers found was 599 ($N = 599$, $n_{\text{EBSCOhost}} = 105$, $n_{\text{Google Scholar}} = 48$, $n_{\text{PSYINDEX}} = 5$, $n_{\text{ResearchRabbit}} = 4$, $n_{\text{Web of Science}} = 437$). Out of these 83 were duplicates (88 were automatically detected by Rayyan, however, 5 were false positives), leaving 516 papers to be screened. During the first screening another 440 papers were excluded. Papers which were excluded did not fit the inclusion criteria, most prominently, they either did not focus on stereotype threat, had the wrong population (e.g., older adults), did not fit the publication type requirements, or did not measure the outcomes of interest - this was assessed using the title, keywords, and abstract. If neither the title, nor the keywords or abstract mentioned enough information to make a decision, the paper was marked as ‘maybe. An example for hypothesis 3 would be, a paper measured working memory but just referred to “the participants” in the abstract, without clarifying that they fit the definition of the academic context. After this first screening, 76 papers remained for the second screening. This second screening was done by looking into the full-text of each paper, here another 44 papers were excluded for the following reasons: wrong focus ($n = 30$), wrong population ($n = 7$), wrong study design ($n = 5$), wrong publication type ($n = 2$) - an overview of this can be found in the PRISMA flowchart in Figure 1. In the end, 32 papers were included in this review, $n = 8$ for hypothesis 1, $n = 9$ for hypothesis 2, and $n = 19$ for hypothesis 3 - some were used for multiple hypotheses. Out of the 516 papers, 319 were excluded for ‘wrong focus’, 57 for

‘wrong population’, 14 for ‘wrong study design’, 12 for wrong publication type, 2 for ‘foreign language’, and 1 for ‘wrong study duration’ (some papers were excluded for multiple reasons). A full list of all papers found and excluded can be found in the GitHub repository.

A template was created to summarise each paper, with two versions completed: one by the author and one by Claude AI. The template was a mixture of the following checklists: CASP systematic review checklist (Critical Appraisal Skills Programme, 2018), Review guidelines for extracting data and quality assessing primary studies in educational research (EPPI-Centre, 2003), Critical appraisal checklist for a systematic review (University of Glasgow, n.d.), and the Newcastle-Ottawa scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses (Wells et al., 2014), which are used to describe studies and assess their quality. Redundant and irrelevant items were eliminated, and the remaining questions were consolidated into a single template. This approach provided a comprehensive overview of the final papers. Based on these summaries, the papers were analysed and the results are presented in the following sections.

RStudio and R packages

The following R packages were used to create this review: R (Version 4.4.1; R Core Team, 2024) and the R-packages *citr* (Version 0.3.2; Aust, 2019), *papaja* (Version 0.1.2.9000; Aust & Barth, 2023), *rmarkdown* (Version 2.27; Xie et al., 2018, 2020), and *tinylabels* (Version 0.2.4; Barth, 2023).

Results

Discussion

References

- Anthropic. (2024). *Claude Ai 3.5 Sonnet*. <https://claude.ai/>.
- Aust, F. (2019). *Citr: 'RStudio' add-in to insert markdown citations*.
<https://github.com/crsh/citr>
- Aust, F., & Barth, M. (2023). *papaja: Prepare reproducible APA journal articles with R Markdown*. <https://github.com/crsh/papaja>
- Barth, M. (2023). *tinylabls: Lightweight variable labels*.
<https://cran.r-project.org/package=tinylabls>
- Critical Appraisal Skills Programme. (2018). CASP Systematic Review Checklist [Organization]. In *CASP - Critical Appraisal Skills Programme*.
<https://casp-uk.net/casp-tools-checklists/>.
- EPPI-Centre. (2003). *Review guidelines for extracting data and quality assessing primary studies in educational research* (Guidelines Version 0.9.7). Social Science Research Unit.
- GitHub, & OpenAi. (2024). *GitHub Copilot*. copilot.github.com.
- Lakens, D. (2022). *Improving Your Statistical Inferences*. Zenodo.
<https://doi.org/10.5281/ZENODO.6409077>
- Ouzzani, M., Hammady, H., Fedorowicz, Z., & Elmagarmid, A. (2016). Rayyan—a web and mobile app for systematic reviews. *Systematic Reviews*, 5(1), 210.
<https://doi.org/10.1186/s13643-016-0384-4>
- Posit team. (2024). *RStudio: Integrated development environment for R* [Manual]. Posit Software, PBC.
- R Core Team. (2024). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.R-project.org/>
- University of Glasgow. (n.d.). *Critical appraisal checklist for a systematic review* [Checklist]. Department of General Practice, University of Glasgow.
- Wells, G., Shea, B., O'Connell, D., Robertson, J., Welch, V., Losos, M., & Tugwell, P.

- (2014). The newcastle-ottawa scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses. *Ottawa Health Research Institute Web Site*, 7.
- Xie, Y., Allaire, J. J., & Grolemond, G. (2018). *R markdown: The definitive guide*. Chapman; Hall/CRC. <https://bookdown.org/yihui/rmarkdown>
- Xie, Y., Dervieux, C., & Riederer, E. (2020). *R markdown cookbook*. Chapman; Hall/CRC. <https://bookdown.org/yihui/rmarkdown-cookbook>

Table 1

Search queries used for the systematic literature review.

Hypothesis	Search Query
H1	("stereotype threat") AND (neural OR neuroimaging OR "functional magnetic resonance imaging" OR fMRI OR electroencephalo* OR EEG OR ERP OR "brain activation" OR amygdala OR "prefrontal cortex" OR "default mode network" OR "salience network") AND (academ* OR education* OR stud* OR learn* OR perform* OR school OR university OR college)
H2	("stereotype threat") AND ("cognitive control" OR "executive function" OR "executive function network" OR "cognitive control network" OR "brain activation" OR "brain activation patterns" OR "cognitive tasks" OR "executive tasks" OR "cognitive assessment" OR "executive assessment") AND (academ* OR education* OR stud* OR learn* OR perform* OR school OR university OR college)
H3	("stereotype threat") AND ("working memory*" OR "processing speed" OR accuracy) AND (academ* OR education* OR stud* OR learn* OR perform* OR school OR university OR college)

Note. The search queries were used in the databases Web of Science, Google Scholar, PSYINDEX, ResearchRabbit, and EBSCOhost. The permalinks to each search used can be found within the GitHub repository.

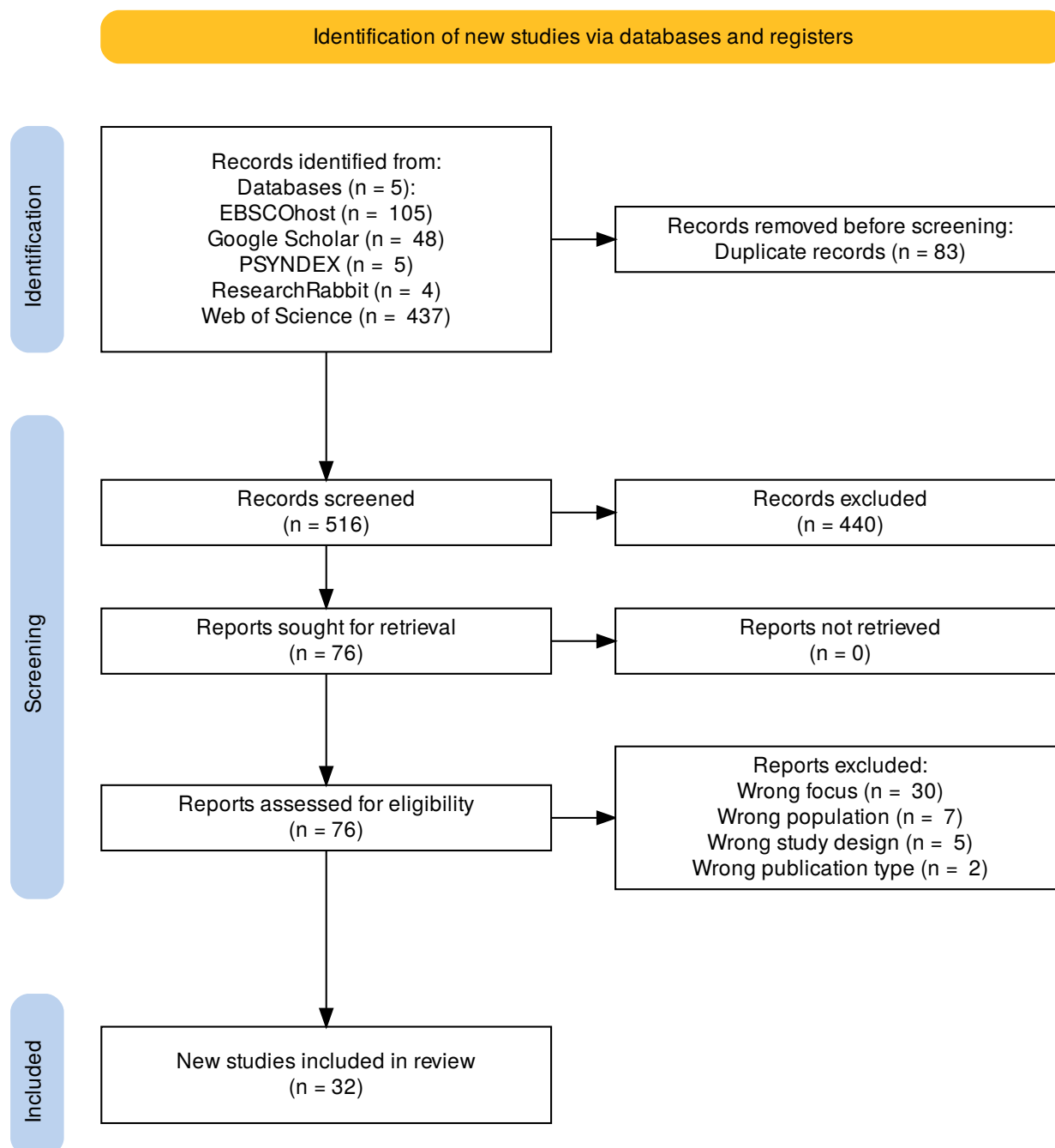


Figure 1

PRISMA flowchart of the screening process.