

**Beispiel für ein Gutachten eines wissenschaftlichen Manuskripts**

Das vorliegende Gutachten wurde zu einem bereits publizierten Artikel gemacht, damit die begutachtete Arbeit allen Interessierten zur Verfügung steht. Das Gutachten wurde so geschrieben, als ob der Artikel als nicht publiziertes Manuskript vorliegt. Das vorliegende Gutachten dient als Beispiel für ein Gutachten zu einem nicht publizierten Manuskript.

Begutachteter Artikel:

Haebig, E., Leonard, L. B., Deevy, P., Schumaker, J., Karpicke, J. D., & Weber, C. (2021). The neural underpinnings of processing newly taught semantic information: The role of retrieval practice. *Journal of Speech, Language, and Hearing Research*, *64*(8), 3195–3211. [https://doi.org/10.1044/2021\\_JSLHR-20-00485](https://doi.org/10.1044/2021_JSLHR-20-00485)

Dieser Artikel ist auf folgenden Webseiten verfügbar:

- [https://pubs.asha.org/doi/10.1044/2021\\_JSLHR-20-00485](https://pubs.asha.org/doi/10.1044/2021_JSLHR-20-00485)
- <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8740735/>

## Gutachten

**Titel des Manuskripts:** The neural underpinnings of processing newly taught semantic information: The role of retrieval practice.

**Autorinnen / Autoren:** Haebig et al.

**Zeitschrift:** Journal of Speech, Language, and Hearing Research

**Zusammenfassung:** Die Autorinnen und Autoren gehen der Frage nach, ob die Wiederherstellung eines Kontexts (*context reinstatement*) eine valide Erklärung für bessere Gedächtnisleistungen bei Abruftraining (*retrieval practice*) sind und welche neurophysiologischen Prozesse dabei involviert sind. Dazu wurden 20 Kinder in zwei Lernbedingungen getestet (Kontextwiederherstellung und Kontrollbedingung). Die Gedächtnisleistung wurde einmal direkt im Anschluss an die Lernphase und nach einer Woche getestet. Zusätzlich wurde ein klassisches Passt / Passt-Nicht Paradigma (*match / mismatch*) mit simultaner Elektroenzephalographie-Messung durchgeführt. Die Resultate zeigen bessere Gedächtnisleistungen für die Kontextwiederherstellungsbedingung im direkten Test und nach einer Woche. Auf neurophysiologischer Ebene wurden in der Kontextwiederherstellungsbedingung frühere (für die N400) und länger (für die LPC) anhaltende Effekte gefunden. Die Autorinnen und Autoren schliessen, dass Kontextwiederherstellung eine valide Erklärung für Abruftraining darstellt.

**Evaluation:** Die vorliegende Arbeit kommt sehr sorgfältig daher. Der Text ist insgesamt sehr verständlich geschrieben. Insbesondere die Einleitung gibt einen guten Überblick über die existierende Literatur zu den untersuchten ereigniskorrelierten Potentialen. Allerdings haben die Autorinnen und Autoren meiner Meinung die elaborative Abrufhypothese (Carpenter, 2009) komplett ausgeklammert. Deswegen bedarf das Manuskript meiner Ansicht einer substantiellen Überarbeitung, bevor ich die Publikation dieser Arbeit empfehlen kann.

### Grössere Schwachpunkte:

- 1) Grösster Schwachpunkt der vorliegenden Arbeit ist, dass sie die elaborative Abrufhypothese (vgl. z.B. Carpenter, 2009) komplett ausklammert. Diese ist aber zentral hinsichtlich der Forschungsfrage, weil die elaborative Abrufhypothese eine mögliche alternative Erklärung für bessere Gedächtnisleistungen bei Abruftraining in der Literatur zu Abruftraining darstellt. Entsprechend muss die elaborative Abrufhypothese bereits in der Einleitung eingeführt und die für die Forschungsfrage relevanten Studien dazu berichtet werden. Zudem sind die vorliegenden Resultate konsistent mit der elaborativen Abrufhypothese. So können zum Beispiel die früheren N400 Effekte in der Kontextwiederherstellungsbedingung hinsichtlich vermehrter automatischer Aktivierungen im semantischen Netzwerk durch schwierigere Lernbedingungen interpretiert werden, was genau der Argumentationslinie der elaborativen Abrufhypothese entspricht (vgl. z.B. Carpenter 2009; Hausman & Rhodes, 2018). Entsprechend müssen die vorliegenden Resultate auch hinsichtlich der elaborativen Abrufhypothese diskutiert werden.
- 2) Die vorliegende Stichprobengrösse ist zu klein für das verwendete Design (vgl. z.B. Brysbaert, 2019). Ich empfehle zwingend zusätzliche Versuchspersonen zu testen, damit sinnvolle statistische Aussagen gemacht werden können. Zudem würde ich

von den Autorinnen und Autoren eine Rechtfertigung im Manuskript hinsichtlich der getesteten Stichprobengrösse erwarten.

- 3) Es wird zwar die Reliabilität der Antwortbewertung berichtet (S. 3199, Abschnitt «Scoring and Reliability»). Allerdings wird die Reliabilität der Aufgabe selbst nicht im Manuskript berichtet. Dies muss zwingend nachgeholt werden, damit sich die Leserinnen und Leser ein Bild von der Zuverlässigkeit der Messungen bzw. Resultate machen können. Da die Chance besteht, dass mit drei Wörtern pro Bedingung (vgl. S. 3198) keine zuverlässigen Messungen möglich sind, würde ich den Autorinnen und Autoren empfehlen zuerst die Reliabilität der vorliegenden Messungen zu bestimmen. Sollte sich jedoch herausstellen, dass die Reliabilität ungenügend ist, würde ich empfehlen eine grössere Stichprobe (vgl. Punkt 2) mit zusätzlichen Wörtern pro Bedingung zu testen. Mir ist bewusst, dass dies mit solch jungen Versuchspersonen kein einfaches Unterfangen ist. Entsprechend würde ich eine Testung über mehrere Blöcke/Tage empfehlen.

#### **Kleinere Schwachpunkte:**

- a) Der erste Absatz (S. 3195) weckt zwar sehr gut das Interesse der Leserschaft. Allerdings fehlen eine Einführung der zentralen Konzepte (wie z.B. retrieval practice), die Forschungslücke und die Forschungsfrage. Dies muss zwingend nachgeholt werden, um das Verständnis hinsichtlich der vorliegenden Arbeit zu verbessern.
- b) Es ist nicht einfach nachzuvollziehen, wie oft und wann ein Wort genau in den verschiedenen Lernphasen präsentiert wurde (S. 3198, Abschnitt «Word Learning Task»). Dieser Teil muss klarer strukturiert werden.
- c) Es ist nicht nachvollziehbar, warum einige Kinder einen anderen nicht-verbale Intelligenztest absolviert haben (S. 3198, Abschnitt «Standardized Assessment»). In der Revision sollte hierfür eine Erklärung im Manuskript gegeben werden.
- d) Im Sinne offener und transparenter Forschung möchte ich den Autorinnen und Autoren die Veröffentlichung der Aufgaben, Stimuli, Rohdaten und Analyseskripte auf Open Science Framework (<https://osf.io/>) bei einer allfälligen Publikation des Artikels empfehlen.

#### **Literatur:**

- Brysbaert, M. (2019). How many participants do we have to include in properly powered experiments? A tutorial of power analysis with reference tables. *Journal of Cognition*, 2(1), 16. <https://doi.org/10.5334/joc.72>
- Carpenter, S. K. (2009). Cue strength as a moderator of the testing effect: The benefits of elaborative retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35(6), 1563–1569. <https://doi.org/10.1037/a0017021>
- Hausman, H., & Rhodes, M. G. (2018). Retrieval activates related words more than presentation. *Memory*, 26(9), 1265–1280. <https://doi.org/10.1080/09658211.2018.1453934>