Study Information

1.   Title

 Spacing effect in learning with text and multimedia

 2.   Authorship

Wach, Stephan, stephan.wach@studium.fernuni-hagen.de

3.   Description and Research Questions

The present study is a replication of a prior work (see master thesis of Petersell, 2019) in which both spacing- and multimedia-effect were investigated, using text-only and multimedia material in spaced and massed learning groups.

Research of repetition has a long tradition, and the so-called spacing effect is one of the most studied phenomena in memory research and highly robust (Carpenter et al., 2012; Cepeda et al., 2006; Latimier et al., 2021, S. 20; Rohrer, 2015; Son & Simon, 2012). It states, that scheduled repetition (spaced learning) leads to better retention performance in a final delayed test than learning only a single period (massed learning). Basically, the retention performance in final tests is dependent from the inter-study-interval (ISI), which is the time between separate learning sessions, and the retention-interval (RI), which is the delay between learning and the final test. In general, spacing effect studies in the past usually dealt with simple learning material. To date, there is no clear statement possible, whether spaced learning is a good strategy for complex learning material like it is for simple learning matter, not to mention reasonable advice for choosing the optimum ISI in dependency of the required RI for complex learning material. Recently there are a few works aiming to fill this gap. Verkoeijen (2007) for example used texts from a non-fiction reading book about the way Hollywood films represent history, and Greving & Richter (2021) investigated distributed learning with multiple complementary texts. "In sum, no benefits of distributed learning vs. massed learning were found, but distributed learning might lower the decrease in learning outcomes over time" stated Greving & Richter. Their work is based on non-repeated complex text material, and they concluded, that more research on distributed learning is necessary for being able to recommend spaced learning for complex material. To contribute findings relating to these issues, complex text material and multimedia material, based on the text material, is used within the present work, adapted from an

experiment of Schüler et al. (2012).

A few considerations about the choice of the ISI-interval of our experiment shall be outlined here: Cepeda et al. (2008) determined optimal values for the ISI depending on the necessary RI, however specifically for simpler materials such as word pairs or vocabulary of a foreign language. They proposed 20% of the RI for an optimal ISI. This study defines the RI to 7 days, so the calculated ISI, according to Cepeda et al., would be 1.4 days. Considering the use of complex material, a shorter ISI seems to be appropriate to the author, to give memory probably more contact points to the first learning session. So, the ISI for this experiment is set to 1 day, instead of two days within the original study. This is the only modified parameter comparing the prior work to the present study.

Multimedia material is used, because this kind of material represents an important part of the learning material common in schools and academic studies. Further to investigate the so-called multimedia effect and possible interaction or additive effects with the spacing effect. The multimedia effect says, that visual information presented in form of text and pictures outmatches the text-only information regarding learning performance (Brünken et al., 2005; Lindner et al., 2018; Scheiter et al., 2018; Schweppe & Rummer, 2016). Using complex text and multimedia material shall contribute to a more ecological setting and lead to findings with value for daily learning routines of learning people.

The combination of the manipulated factors of this experiment leads to four learning groups: massed/text-only, massed/multimedia, spaced/text-only and spaced/multimedia. Data analysis of the previous master thesis couldn't identify any of the hypothesized effects, which were main spacing effect, main multimedia effect, and additive spacing-multimedia effect on long-term retention in a final test. Within this final test three types of questions were used to generate the dependent variables, measuring the performance on free recall, open questions and transfer questions. Further analysis of single effects of the master thesis data showed several interesting findings. First, significant single effects for both spacing and multimedia could be found for free recall and open questions. Second, the data showed a small antagonistic interaction of spaced and multimedia learning for free recall and transfer questions, however not significant. The performance in spaced/multimedia groups for these two kind of questions was lower than the performance in spaced/text-only or multimedia-massed groups. Overall, the performance of spaced/multimedia groups was comparable to the

performance of massed/text-only groups. For open questions, the performances of spaced/text-only groups, massed/multimedia groups and spaced/multimedia groups were almost similar and for each group significantly higher than for massed/text-only groups.

The present study is launched to explore, whether these findings can be replicated and to possibly expand the insights relating to choosing an appropriate ISI for a given RI.

Study Information: Hypotheses

Hypothesis 1 Spacing effect:

Hypothesis 1.1: Spaced learning with text-only material leads to a higher retention performance with free recall, open and transfer questions in the final test, compared to massed learning of monomedia texts.

Hypothesis 1.2: Spaced learning with text-only material leads to a similar retention performance with free recall, open and transfer questions in the final test, compared to massed learning of monomedia texts.

Hypothesis 2 Multimedia effect:

Hypothesis 2.1: Learning with multimedia material leads to a higher retention performance with free recall, open and transfer questions in the final test, than learning with monomedia text in the massed learning group.

Hypothesis 2.2: Learning with multimedia material leads to a similar retention performance with free recall, open and transfer questions in the final test, as learning with monomedia text in the massed learning group.

Hypothesis 3 Possible interaction between the spacing and multimedia effect:

Hypothesis 3.1: Combining spaced learning and multimedia material leads to a lower or similar retention performance with free recall, open and transfer questions in the final test, than spaced learning alone does.

Hypothesis 3.2: Combining spaced learning and multimedia material leads to a lower or similar retention performance with free recall, open and transfer questions in the final test, than multimedia alone does.

Design Plan

1. **Study Type**

   Experiment

2. **Blinding**

- For studies that involve human subjects, they will not know the treatment group to which they have been assigned.
- Personnel who interact directly with the study subjects (either human or non-human subjects) will not be aware of the assigned treatments ("double-blind").

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

- Participants are not informed about the study design, and they do not receive any information in advance which of the 4 counterbalancing groups they will belong to.

4. **Study Design**

- This study includes a two-factorial design, with the between-subject factors being *study practice* (spaced vs. massed) and *learning material* (text-only vs. multimedia).

- The main dependent variable is free recall performance (i.e., proportion of correctly recalled idea units), which is measured by a final test after 7 days. Additionally, for further investigation, multiple choice questions (i.e., choosing 1 correct answer out of 4 options), open questions (i.e., free answers to specific questions) and problem-solving questions (i.e., testing transfer knowledge) are part of the test.

- Design-Plan:

Practice Type	Learning Material	Session 1 — Learning Phase			ISI	Session 2 — Learning Phase				RI	Session 3 — Final Test Phase				
Massed	Multimedia					Information and Pre-Test	Study-Phase 1	Study-Phase 2 - Restudy	Measures 1+2	7 Days	Judgement of Learning (JOL)	Test 1 Free recall Question	Test 2 Mixed Questions (23): Open, MC- and Transfer Questions	Paper-Folding-Test (PFT)	Debriefing and Meaures 3
	Text-only														
Spaced	Multimedia	Information and Pre-Test	Study-Phase 1	Measures 1	1 Day	Study-Phase 2 - Restudy			Measures 2						
	Text-only														

5. Randomisation

Practice groups and learning groups will be assigned randomized across participants.

Sampling Plan

1. Existing Data

Registration prior to creation of data

Registration prior to any human observation of the data

Registration prior to accessing the data

Registration prior to analysis of the data

Registration following analysis of the data

2. Explanation of Existing Data

3. Data Collection Procedures

We plan to collect data from participants aged 18 and older. Undergraduate students will be tested from the "FernUniversität in Hagen", "Bielefeld University" and "Humboldt Universität zu Berlin" in spring 2022 during the time period from 24th of March 2022 to 30th of April 2022. Participants will be recruited through the "Virtuelles

Labor" (research portal for psychology students of the FernUniversität in Hagen), Facebook-groups, other social media and private contacts.

We use the Gorilla Experiment Builder (www.gorilla.sc, Anwyl-Irvine et al., 2020) to create and host the study.

4. Sample Size

According to the a priori power analysis, we aim to get a total number of $N = 180$ participants, 45 for each group.

5. Sample Size Rationale

Calculation a priori with G*Power (Faul et al., 2007, 2009) and the assumptions F-Test ANOVA (fixed effects, special, main effects an interactions) with four groups, $\alpha = .05$, pursued Power = 0.9, a medium effect size of $f = 0,25$, numerator $df = 3$. This results in a total sample size of 231, for 4 groups, 58 participants per group (rounded).

6. Stopping Rule

The data will be collected from the 24th of March 2022 until the 30th of April 2022 or until we have reached 232 participants that have successfully participated in both study sessions.

Variables

1. Manipulated Variables

In the present study, two independent variables are manipulated:

1. *Study practice* (spaced vs. massed) and

2. Learning material (text-only vs. multimedia).

Participants are randomly assigned to the counterbalancing factor of whether spaced/massed learning or text-only/multimedia occurs in learning sessions.

2. Measured Variables

Measured variables

- *Recall performance* (i.e., proportion of correctly recalled idea units from the key-term definitions) is assessed by means of a cued-recall test after 7 days.

- *Cognitive effort* is assessed by a short one-item scale (Paas et al., 2008) after every study phase. Answers are provided on a 7-point Likert-type scale from 1 (*very low*) to 7 (*very high*; (Schmeck et al., 2015)).
- *Prior knowledge* is assessed for all participants by a test before the first study phase using a scale (6 MC-Questions).
- *Demographic variables*. Participants report age, gender, native language, and educational attainment level at the end of Session 2.

Measured variables for further exploratory analyses

- *Disturbance* during the study, assessed at the end of every session, using a Likert-type scale of possible integer values between 0 (*very little*) and 10 (*very much*).
- *Concentration* during the study, assessed at the end of every session, using Likert-type scale of possible integer values 0 (*very bad*) and 10 (*very good*).
- *Difficulties* during the study, assessed at the end of every session (*yes, vs. no*).
- *Previously participated in a FET study*, assessed at the end of the last session (*same study*, *similar study*, vs. *no*).
- *Study commitment* as assessed by two statements: *I seriously participated and processed the task in this study* versus *I did not seriously participate and process the tasks in this study.*
- *Use of external aids* (e.g., writing words on a piece of paper) during the study, assessed at the end of Sessions 1 and 2 (*yes, no,* vs. *no answer*).
- *Intrinsic motivation*, using the short version of the Intrinsic Motivation Inventory (IMI) scale (Ryan & Deci, 2000). Specifically, we used three questions of the enjoyment subset of the IMI scale, from 1 (*not at all*) to 7 (*exactly*).
- *Performance satisfaction* ("I am satisfied with my performance in repeating the text") *(not at all)* to 7 *(exactly)*.
- *Skilful approach* (When repeating the text, I made a skilful approach) from 1 *(not at all*) to 7 *(exactly)*.
- *Interest in the text* (after each study phase) for all participants using a scale 0 (*very little*) to 9 (*very much*).
- *Joy of reading*, assessed at the end of session 2 (*sehr ungern, eher ungern, eher gern, sehr gern*).

- Participants *reading time* in the week given in hours.
- Readability of study materials s, assessed at the end of all the learning sessions (*yes/no/slight problems*).
- Skill of using computer keyboard, assessed at the end of the last session using a scale from 0 (*very little*) to 10 (*very much*).
- Participation in an exam within the last 14 days, assessed at the end of the last session (*no, 1, 2 or more*).
- Spatial Reasoning Ability (Paper Folding Test (PFT)).

These can be modelled as either mediators, moderators, or criterion variables, dependent on the specific research question, or they are used as exclusion criterion.

3. **Indices**

Analysis Plan

1. **Statistical Model**

The statistical model is a balanced, two-way ANOVA with the grouping variables "repetition" (levels spaced vs. massed) and "study material" (levels text-only vs. multimedia). The dependent variables are the retention performances measured after 7 days by a final knowledge-test, with three different forms of questions: free recall, open questions, transfer questions.

2. Transformations

- Not planned.

3. Inference criteria

Standard alpha level of $\alpha = .05$.

4. Explanatory notes to hypotheses

The hypotheses are formulated in two ways (numbered 1.1 and 1.2, 2.1 and 2.2, 3.1 and 3.2). For the first part of every hypothesis a traditional null-hypothesis significance test will be performed, for the second part an equivalence test (see Lakens et al., 2018).

For the equivalence test of hypotheses 1 and 2 (regarding spacing and multimedia effect), the SESOI (smallest effect size of interest) is set to $d = 0.25$ (small effect), and for hypothesis 3 (interaction of spacing and multimedia effect) to $d = 0.2$ (small effect). This is based on research of past experiments relating spacing and multimedia effect and due to the results of the replicated study (Petersell, 2019).

5. Data exclusion

The following exclusion criteria are applied, based on a prior online data collection:

1. Incomplete participation (removed immediately).
2. Participated in the same study (participation_already).
3. Used external resources (external.help_poststudy, external.help_spaced.poststudy2).
4. Participants will be excluded from the final analyses, if they had a low recall performance ($x_i < M_i - 2.5 *MAD$) on the recall test or at least the proportion of correctly recalled information being less than .1, with $x_i = n$ original observations, M = Median, MAD = Mean Absolute Deviation $= b* M_i (|x_i - M_j (x_j)|)$, and $b = 1.4826$ (cf. Rousseeuw & Croux, 1993), and fulfill one of the following conditions below (a-d):

    a) Participants are not fluent in German (i.e., they are not native in German, or moved to Germany or a German-speaking country before they were 16).

b) Participants report to be either too distracted from their environment (disturbed_poststudy OR disturbed_poststudy1 OR disturbed_spaced.poststudy2 with 8 OR more on a 10-point Likert response scale) AND they could not sufficiently focus on the experiment (concentration_poststudy or concentration_poststudy1 or concentration_spaced.poststudy2 with 3 or less on a 10-point Likert response scale).

c) Participants chose the items that they did not seriously participate and process the tasks in the study.

d) Participants provide in the free comment box that they were not motivated or did not participate seriously.

Furthermore, participants will be excluded if they
- indicate that they previously participated in the same study (including the same learning materials), or
- recall odd contents not related to any presented text sections or similar responses showing that they were not motivated or did not participate seriously, or
- indicate that they used external resources during the experiment (e.g., taking notes, writing down the learned items).

To achieve the highest data quality possible, we will further carefully monitor suspicious behavior in our data after the preregistration that suggests participants were not taking the experiment seriously. This may result in additional, non preregistered exclusion criteria that we would transparently describe in potential scientific publications and would subsequently include as a preregistered exclusion criterion in potential follow-up experiments.

6. **Missing data**

7. **Exploratory analysis**

Other

This preregistration intentionally builds on prior preregistrations on the spacing effect and the multimedia effect: To increase comparability and transparency between

preregistrations, many aspects are kept identical to previous preregistrations, and only critical deviations have been changed in the current preregistration.

References

Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods*, *52*(1), 388–407. https://doi.org/10/gfz97m

Brünken, R., Seufert, T., & Zander, S. (2005). Förderung der Kohärenzbildung beim Lernen mit multiplen Repräsentationen: Fostering Coherence Formation in Learning with Multiple Representations. *Zeitschrift für Pädagogische Psychologie*, *19*(1/2), 61–75. https://doi.org/10/b37nqh

Carpenter, S. K., Cepeda, N. J., Rohrer, D., Kang, S. H. K., & Pashler, H. (2012). Using Spacing to Enhance Diverse Forms of Learning: Review of Recent Research and Implications for Instruction. *Educational Psychology Review*, *24*(3), 369–378. https://doi.org/10/gc7mhj

Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T., & Rohrer, D. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin*, *132*(3), 354–380. https://doi.org/10/cnpvd9

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), 1149–1160. https://doi.org/10/b22kn7

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10/bxjdcg

Lakens, D., Scheel, A. M., & Isager, P. M. (2018). *Equivalence Testing for Psychological Research: A Tutorial*. 11.

Latimier, A., Peyre, H., & Ramus, F. (2021). A Meta-Analytic Review of the Benefit of Spacing out Retrieval Practice Episodes on Retention. *Educational Psychology Review*, *33*(3), 959–987. https://doi.org/10/gjqf67

Lindner, M. A., Eitel, A., Barenthien, J., & Köller, O. (2018). An integrative study on learning and testing with multimedia: Effects on students' performance and metacognition. *Learning and Instruction*, 101100.

https://doi.org/10.1016/j.learninstruc.2018.01.002

Paas, F., Ayres, P., & Pachman, M. (o. J.). *Assessment of Cognitive Load in Multimedia Learning*. 27.

Petersell, G. (2019). *Spacing Effect in Learning with Graphs* [Unpublished master's thesis]. FernUniversität in Hagen.

Rohrer, D. (2015). Student Instruction Should Be Distributed Over Long Time Periods. *Educational Psychology Review*, *27*(4), 635–643. https://doi.org/10.1007/s10648-015-9332-4

Scheiter, K., Richter, J., & Renkl, A. (2018). Multimediales Lernen: Lehren und Lernen mit Texten und Bildern. In H. Niegemann & A. Weinberger (Hrsg.), *Lernen mit Bildungstechnologien*. Springer-Verlag GmbH Deutschland, ein Teil von Springer Nature 2018.

Schmeck, A., Opfermann, M., van Gog, T., Paas, F., & Leutner, D. (2015). Measuring cognitive load with subjective rating scales during problem solving: Differences between immediate and delayed ratings. *Instructional Science*, *43*(1), 93–114. https://doi.org/10.1007/s11251-014-9328-3

Schweppe, J., & Rummer, R. (2016). Integrating written text and graphics as a desirable difficulty in long-term multimedia learning. *Computers in Human Behavior*, *60*, 131–137. https://doi.org/10.1016/j.chb.2016.02.035

Son, L. K., & Simon, D. A. (2012). Distributed Learning: Data, Metacognition, and Educational Implications. *Educational Psychology Review*, *24*(3), 379–399. https://doi.org/10/gfsdk3