

Segmenting in Multimedia Learning

Shreya Sheel

UC San Diego Cognitive Science Honors Program 2020

Committee: Dr. Drew Walker, Dr. Emma Geller, Rachel Bristol

Abstract

In multimedia learning, segmentation is offered as a technique to reduce cognitive load demands. Based on the cognitive theory of multimedia learning, cognitive overload occurs when the learner's intended cognitive processing exceeds the learner's available cognitive capacity. By breaking down presentations into bite-size chunks, segmenting gives learners the time and capacity to process and organize selected information (Mayer, 2003). Although segmentation is recommended as a method to reduce cognitive overload, the optimal amount of segments to achieve this is still unclear. Our study is an attempt to investigate how to implement segmentation as a technique to improve learning outcomes and manage cognitive overload. We explore the effects of various degrees of segmentation by manipulating the number of segments (1, 5, 12, and 25) to determine whether increased segmentation improves retention and transfer test performance. Additionally, we test for cognitive load measures and examine if segmentation plays a role in reducing intrinsic load and germane loads, which closely map to the essential processing demands learners face when making sense of complex material and creating mental models. Results suggest that segmentation did not increase retention and transfer scores and also did not reduce intrinsic and germane loads. This poses an interesting discussion in how the complexity and length of lessons play a role in how segmentation can be used as an effective load-reduction method.

Introduction

Multimedia instruction has become increasingly popular in classroom and online settings as more digital tools and resources have emerged. With the rise of online courses and remote learning, it has become even more important to understand Richard Mayer's (2005) Cognitive Theory of Multimedia Learning (CTML) which touches on best practices in creating materials containing both pictures and words in order to achieve meaningful learning. CTML is based on three assumptions. The first is the dual channel assumption: the idea that separate channels in working memory are used to process visual and verbal information. The second assumption is that our working memory only has a limited amount of processing capacity available in these visual and verbal channels. The final assumption is that learning requires active cognitive processing in the verbal and visual channels as opposed to passive learning in order for meaningful learning to occur. In other words, a learner must intentionally choose relevant information to take into the working memory and engage with it to achieve better understanding. Meaningful learning means deep understanding of material as determined by recognizing important parts of the material, creating coherent mental models, and integrating it with relevant existing knowledge (Mayer, 2003).

According to Mayer (2003), the main challenge in designing multimedia learning materials is avoiding cognitive overload, where the learner's intended cognitive processing demands exceed the learner's cognitive capacity. These cognitive demands are sometimes referred to as loads (Sweller, 1999) and other times referred to as processing, depending on the researcher. However, since their intervention methods to measure the properties of the learning materials or situations are similar, these terms can roughly be used interchangeably. Cognitive demands are distinguished into three types. The first is germane load, or essential processing,

which refers to the cognitive processes involved in making sense of presented material to select, organize, and integrate words and images. Extraneous load, or incidental processing, is the cognitive processes that are non-essential in understanding presented material. Lastly, intrinsic load, or representational holding, is the cognitive processing involved in storing verbal or visual representations in the working memory.

In this paper, we are specifically interested in looking at what Mayer (2003) classifies as Type 2 overload, or when visual and auditory channels in working memory are overloaded with essential processing demands. If information is presented too fast or is too content-dense, learners may not have enough time to develop coherent mental models in organizing the presented words and images. Sweller (1999) refers to this presentation of material as a situation with high-intrinsic load, where the material is high in complexity. In these situations, complexity is determined by the number of elements and the relations between them. When the material is too complex, the demands of essential processing can overwhelm the learner (Mayer, 2008). In order to manage this cognitive overload from complexity, Mayer offers segmentation as an essential processing load-reduction method. Segmentation provides learner control over the pace of instruction and allows the learner to fully represent each part of a system before moving on to the next (Mayer, 2008). Although the material's content cannot reduce in complexity, segmentation allows learners to process and connect information from the presentation one part at a time and build component models, creating for more opportunities in deeper levels of learning (Mayer, 2001).

Mayer tests this theory with a 140 second lesson on lightning formation, breaking up the video into 16 segments of about 8 to 10 seconds each. Learners had control over when they wanted to proceed to the next segment. Results illustrated that participants who received the

segmented presentation performed better on subsequent tests of problem-solving transfer than did participants who received a continuous presentation (Mayer & Chandler, 2001, Experiment 2; Mayer, Dow, & Mayer, 2003, Experiments 2a and 2b). Another study further explored this principle by testing the effect of the degree of segmentation, as determined by the number of segments, in affecting students' recall and application of presented information. The study used a 9 minute historical lesson, broken up into either 1, 7, 14, or 28 segments. The lesson was segmented into conceptually coherent increments. Participants had control over time spent between segments before moving forward but could not control the content pace or stop/rewind. They were then tested on how well they executed a writing and reading strategy to interpret primary sources in history against a writing rubric. Results indicated that increased segmentation facilitated recall and application (Doolittle et al, 2015).

Despite evidence of the benefits of segmentation in the studies cited above, previous unpublished work in our lab has failed to replicate the benefit of segmentation. There are a number of differences between this past work and Mayer's studies, including video length, lesson topic, number of segments, and segment length. The purpose of this study is to further explore the segmentation effect in order to better understand when segmentation will or *will not* lead to improved learning. The focus of the current study is a manipulation of the number and length of video segments. If Mayer's studies succeeded because his lesson contained many, short segments, then we should expect that our participants will learn more in conditions with more segments that are shorter. If the benefits of segmentation are unrelated to the number or length of segments, then we should not expect to see any learning differences across our conditions.

Our study aims to apply the type of approach from the Doolittle study in terms of degrees of segmentation but instead use a video lesson similar to Mayer's, in that participants learn about

a scientific system rather than apply a writing strategy. Similar to the above studies, meaningful learning will be measured through retention and transfer tests. Retention tests ask learners to recall or recognize parts of the presented material while transfer tests ask learners to apply what was learned in a new situation (Mayer, 2008). Additionally, we assess essential processing demands through germane and intrinsic load ratings, as these are traditional load measures that most closely measure creating meaning around the material and interpreting complexity, respectively.

Hypothesis

Based on results of past studies, which suggest segmentation positively influences learning, we hypothesized that participants who watched videos with increased degrees of segmentation would achieve higher scores on retention and transfer tests. Participants in conditions with higher degrees of segmentation would also rate their intrinsic loads lower and germane loads higher.

Methods

Participants

A total of 234 undergraduate students (57 males, 175 females, 1 other, and 1 undisclosed) with a mean age of 20.44 years (2 declined to state) from the University of California, San Diego participated in the study. The sample population was obtained using UCSD SONA Systems, an online recruiting system through which undergraduate students can earn course credit for participating in research studies. Participants were randomly assigned to one of 4 between-subjects conditions: 1 segment ($n = 60$), 5 segments ($n = 58$), 12 segments ($n = 57$), or 25 segments ($n = 59$).

Materials

All participants completed the study on desktop iMacs in a laboratory located at the University of California, San Diego. The study was administered via a Qualtrics survey and videos were embedded through YouTube.

Video Lesson

Participants watched a short video lesson called “How Touchscreens Work in Simple Words” created by BrightSide. The lesson covered several topics including the different types of touchscreens, their applications, and the benefits/drawbacks of each. The video (with credits removed) was 9 minutes long and was professionally produced for a YouTube channel called BrightSide.

Segments

For the experimental conditions, the video was divided into 5, 12, and 25 segments with the 5 segments condition averaging a length of 112.4 seconds (range = 28 - 154 seconds), the 12 segments condition averaging a length of 46.33 seconds (range = 26 - 96 seconds), and 25 segments condition averaging a length of 21.68 seconds (range = 20- 30 seconds). We identified segments by extracting different topics within the lesson and divided by transitions between concepts. Each segment was displayed on a separate page of the Qualtrics survey. Participants had to click a continue button at the bottom of the page after watching each segment in order to advance to the next.

Prior Knowledge

In order to assess prior knowledge, participants were asked “How much do you know about how touchscreens work?” and to rate their response on a scale from indicating their knowledge 1 (a little) to 5 (a great deal).The average rating across all participants was 3.50 (SD

=1.187). They were then asked to demonstrate their prior knowledge rating by answering the question, “How many different kinds of touchscreens are there? List all of the types you know below.” Out of all 234 participants, not a single participant named the correct number of touchscreens. Thus, we are confident that none of the participants had extensive prior knowledge about the content of this lesson.

Cognitive Load

In order to measure cognitive load, we used the Paas scale (Leppink et al, 2013). Participants were asked to self-evaluate their learning experience by rating their opinions on a scale from 0 (not at all the case) to 10 (completely the case). We assessed intrinsic load by asking participants to indicate the extent to which they agree/disagree with 4 statements such as, “The vocabulary used in the video was very complex”. Extraneous load was evaluated through 4 statements like “The explanations in the video were full of unclear language”. Germane load was determined through 3 statements such as “The video really enhanced my understanding of how touchscreens work”. The total score for each cognitive load type was calculated by taking the average rating for items on each subscale.

Retention Test

The retention test was designed to measure what participants remembered from the lesson. It consisted of 2 open responses and 12 multiple choice questions. The open response questions were compared to a predetermined list of answers, containing anywhere from 11 to 12 possibilities, and the multiple choice questions were worth 1 point each.

The open response questions were intended to test participants in recollection of certain information while multiple choice focused on recognition. Performance on multiple choice and open response questions were analyzed separately.

The first open response question asked participants to describe how resistive and capacitive touchscreens are different in much detail as possible. We generated a list of the core ideas presented in the lesson that were related to this question, and participants earned one point for each idea they included in their answer. The second open response question asked participants to describe how three other kinds of touchscreen mentioned work and how they are different from each other. Answers were scored according to the number of functions and differentiations named. The multiple-choice questions tested participants' knowledge on types of touchscreens, their mechanisms, and common applications.

Transfer Test

The transfer test was designed to measure if participants could apply information from the lesson to a different situation. The test consisted of 4 open response questions that were compared to a predetermined list of answers, containing anywhere from 1 to 3 possibilities. Participants earned a point for every reasonable idea they included in their answer, as determined by a list of possible answers that was generated ahead of time. The transfer score was calculated by taking the proportion of points earned out of total points possible across all 4 questions.

The first open response question asked participants to explain why or why not a stylus would work on a capacitive screen. The second question asked to troubleshoot why your phone wasn't responding to your touch if you just got out of the pool. The third question asked to design gloves that would respond to your smartphone and describe what they would need to work and why. Answers were scored as either correct or incorrect with a few different valid answers. The fourth question asked to predict how the functionality of your smartphone would change if it was resistive rather than capacitive and to list drawbacks of this choice.

Procedure

Participants completed the study in a laboratory on UCSD's campus. Each participant encountered the same introductory instructions stating "you will watch a brief lesson on how touchscreens work and then take a short test to measure what you have learned." They then answered a few demographic questions including age, gender, whether or not English was their native language, and what age they began learning English. Participants then rated their prior knowledge and were assigned to one of the four segmented conditions. Participants in all four conditions were required to watch the same lesson. After each segment, a continue button appeared at the bottom of the screen to allow them to proceed to the next segment. Each page of the Qualtrics survey was timed, forcing participants to stay on the page for the entirety of the segment and preventing them from speeding up the video's content to move onto the next portion. Once they completed watching all segments, participants answered a few questions giving feedback about their impressions of the lesson, namely if they found any parts of the lesson confusing. Afterwards, they completed the cognitive load measure ratings. . They then played Tetris for 1 minute as a brief distractor task before proceeding to the retention, transfer, and multiple choice tests. After completing the tests, participants were debriefed and thanked for their participation.

Results

Performance was evaluated and compared to determine if segmentation provided higher scores for those in the experimental conditions as compared to the control condition. In order to address the issue of consistency in evaluating participant answers, two raters scored retention open response and transfer open response. The correlation between the two raters computed for retention open response and transfer open response was 0.79 and 0.80, respectively. In cases

where the two raters gave different scores, we chose to include the lower of the two scores in the analysis. Exploratory analyses revealed that the results did not depend on whether we used just one rater's scores or the lower of the two, so we will report results from the lowest rating analysis.

Prior to analyzing the data, 3 participants were excluded because they were unable to complete the experiment due to technical difficulties. Across all participants, the average score for retention open response questions was 0.23 (SD= 0.15), for retention multiple choice was 0.63 (SD = 0.14), and for transfer open response was 0.63 (SD = 0.22). The average score across all participants for intrinsic load was 3.27 (SD = 1.64) and for germane load was 8.21 (SD = 2.03). Averages describing participants' learning experiences as related to their mental effort, understanding level, concentration level, topic interest, and engagement level were also measured (see Appendix A).

Additionally, we checked if prior knowledge rating was related to participants' test performances. There were no significant differences in prior knowledge ratings across conditions ($p > 0.05$). However, we found that prior knowledge rating was predictive of test performance for retention open response ($F(1,232) = 6.722, p = 0.010$), retention multiple choice ($F(1,232) = 6.575, p = 0.011$), and transfer open response ($F(1,232) = 3.890, p = 0.0498$). For this reason, prior knowledge rating was included as a covariate in further analyses.

Following these pre-test checks, we analyzed the effect of the degree of segmentation on each component of the final test (retention open response, retention multiple choice, transfer open response), using a one-way ANCOVA with prior knowledge rating as a covariate. There was no significant effect of the degree of segmentation on test performance for any test type (all

$p > 0.05$). However, a general increasing trend in the average of scores as segments increased is seen. Table 1 lists specific values.

Table 1

Mean Proportion Correct of Retention Multiple Choice, Retention Open Response, and Transfer Open Response Questions for Each Segmented Condition

	<u>1 segment</u>	<u>5 segments</u>	<u>12 segments</u>	<u>25 segments</u>
Retention Multiple Choice	0.61 (0.02)	0.64 (0.03)	0.62 (0.03)	0.60 (0.03)
Retention Open Response	0.22 (0.02)	0.22 (0.03)	0.23 (0.03)	0.23 (0.03)
Transfer Open Response	0.62 (0.03)	0.64 (0.04)	0.61 (0.04)	0.64 (0.04)

Note: Standard deviations are in parentheses.

To determine whether segmentation led to decreased essential processing demands, we then analyzed the effect of condition on intrinsic load and germane load using separate one-way ANOVAs. Because we did not find a significant relationship between cognitive load and prior knowledge ($p > 0.05$), we did not include it as a covariate in these analyses. We found that intrinsic cognitive load does not significantly differ across conditions ($F(3,230) = 1.601$, $p > 0.05$). However, there was a general trend for intrinsic load to be rated lower as the number of segments increased, as illustrated in Figure 1.

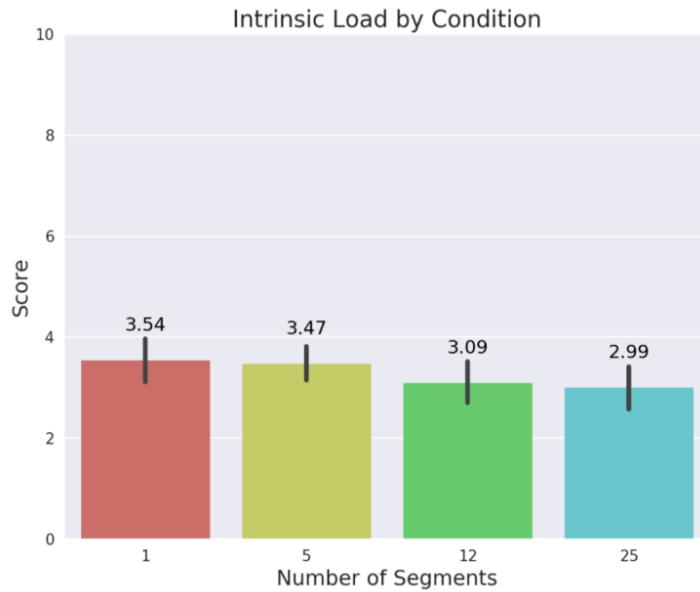


Figure 1: Intrinsic cognitive load score (out of 10) for each segmented condition. Error bars represent standard errors of the mean.

Similarly, we found that germane cognitive load does not significantly differ across conditions ($F(3,230) = 1.268 \text{ p}> 0.05$). There was also a general trend of lower germane load ratings as the number of segments increased, as illustrated in Figure 2.

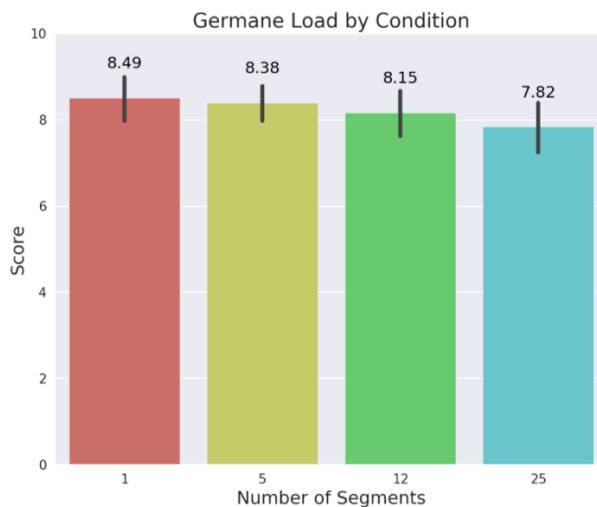


Figure 2: Germane cognitive load score (out of 10) for each segmented condition. Error bars represent standard errors of the mean.

In an exploratory analysis, we investigated whether the effect of condition on test performance might depend on whether the participants are native English speakers or not. We ran a 2x4 between-subjects ANOVA with native English status and condition as predictors, and each test type as a separate outcome variable. Although there was no significant interaction between segmenting and native English status for any test component [all $p > 0.05$], some interesting patterns were found in the retention open response scores as displayed in Figure 3 below. Native speakers performed the best on this test when they saw 12 segments, but non-native speakers performed best in the 25 segments condition. This difference was not significant but indicates that language status may play an important role in the effectiveness of segmenting.



Figure 3: Proportion score of retention open response points earned in each segmented condition for each speaker type: non-native English speakers (red) and native English speakers (yellow).

We ran additional exploratory analyses to determine whether other types of cognitive load predict test performance. We ran separate multiple regression analyses for each test type with intrinsic, extraneous, and germane load as simultaneous predictors. The only significant findings were (1) that germane load was a significant positive predictor of transfer open response scores ($F(1,230)=5.152$, $p=0.0241$, Std Beta=0.153), but not on either retention test (all $p>0.05$), and (2) that extraneous load was a significant negative predictor of retention open response ($F(1,230)=3.981$, $p=0.0472$, Std Beta= -1.50), but not on retention multiple choice or transfer open response (all $p>0.05$). Intrinsic load was not a significant predictor of performance on any test (all $p>0.05$). Higher germane load predicted higher transfer scores, while higher extraneous load predicted lower retention scores.

Discussion

Overall, we found that our results were inconsistent with past research that segmenting would improve retention and transfer test performance. Segmentation also did not have an effect in reducing any cognitive load type. However, germane load positively predicts transfer scores while extraneous load negatively predicts retention scores. Intrinsic load was unrelated to any test performance.

One potential reason as to why segmenting did not have an effect on test performance may be around how participants spent time in between segments. We did not monitor what participants did during pauses; rather, we only know how much time they spent on a segment before moving onto the next. Therefore, it's not possible to determine whether test performance depended on participants' activity in between segments. Additionally, although we told participants they were not allowed to speed up videos and forced-timed them to stay on the page

for each segment's duration, we could not actively enforce them to engage with the lesson in the way we envisioned.

Another potential reason as to why segmenting did not have an effect on test performance could be because of the lesson's complexity and length. Participants self-evaluated their intrinsic loads as fairly noncomplex with scores averaging about 3.23 out of 10 across all conditions. Past studies have not specifically measured intrinsic load ratings as a justification for choosing the tested lessons; however, this poses an interesting nuance that segmentation may only reduce essential processing if the material is perceived as complex. Segmenting videos could perhaps only be beneficial in improving learning outcomes if the lesson is overwhelming to begin with, when essential processing is overloaded. On the contrary, if segmentation is supposed to work on any kind of lesson, no matter the complexity, then it is unclear why we did not find significant results. The video duration and subject could also contribute to the perceived complexity and nonsignificant results. Past studies with similar video durations tested participants on material less related to systems, and more related to interpreting sources based on evidence. Furthermore, the study that did actually test participants on their system knowledge only used a video that was less than 3 minutes with each segment averaging 8 to 10 seconds. Thus, the effect of degree of segmentation may differ depending on the material's subject matter and duration.

During our study, we also encountered a few distracting limitations that participants explicitly reported in their posttest feedback that disrupted their concentration. Some of these limitations were due to our reliance on YouTube. Qualtrics does not allow clips larger than 100 MB to be directly imported onto its platform; most of our clips were larger. Therefore, we used YouTube to upload the clips before embedding them within Qualtrics. However, using YouTube brought up a number of issues. We had no control over participants potentially choosing closed

captioning as well as changing the speed of the video playback. YouTube also displays a screen of “recommended videos” with various related topics at the end of each clip which some participants reported as distracting. Another distracting limitation was due to construction right outside the lab space. Loud drilling noises occurred in the middle of videos during some participant studies, which made hearing the videos hard and overall difficult to focus.

Through our exploratory analyses results, we found that participant speaker status may play a role in test performance. We noticed a trend in non-native speakers performing best in the condition with the most segments, while native speakers performed best in a condition with fewer segments. Upon further investigation, we also found that nonnative speakers generally rated 25 segments lower in their perceived intrinsic load than native speakers, suggesting that segmenting may reduce the perceived complexity of the lesson. However, this would need to be replicated in another study with a much larger sample before we could draw any reliable conclusions about whether segmentation actually helps and whether the way it helps is different for native and non-native speakers. A potential future direction could consider participants’ language capabilities more closely in terms of its role in the effectiveness of segmentation interventions.

Future studies around the segmentation principle should explore types of lessons used and identify topics that participants report more overwhelming in complexity initially. These studies should also track participant activity in between segments and experiment with different ways participants use their time such as engaging in review questions or sitting through fixed pauses. Future research could also consider different numbers of segments. We explored only four possible segmentation conditions of infinite possibilities, although we do not expect the benefits of segmentation to continue into infinitely small units. There may be a specific point at

which further segmentation is no longer effective. Future studies could explore where this cutoff could be. Finally, there may be a relationship between the amount of content within each segment. The content contained in each segment may play a larger role in learners creating mental models than simply the number of segments independently.

References

- Doolittle, P. E., Bryant, L. H., & Chittum, J. R. (2015). Effects of degree of segmentation and learner disposition on multimedia learning. *British Journal of Educational Technology*, 46(6), 1333-1343.
- Kalyuga, S., Chandler, P., & Sweller, J. (1999). Managing split-attention and redundancy in multimedia instruction. *Applied Cognitive Psychology*, 13, 351–371.
- Leppink, J., Paas, F., Van der Vleuten, C. P., Van Gog, T., & Van Merriënboer, J. J. (2013). Development of an instrument for measuring different types of cognitive load. *Behavior research methods*, 45(4), 1058-1072.
- Mayer, R. E., & Chandler, P. (2001). When learning is just a click away: Does simple user interaction foster deeper understanding of multimedia messages? *Journal of Educational Psychology*, 93(2), 390–397.
- Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational psychologist*, 38(1), 43-52.
- Mayer, R. E. (2008). Applying the science of learning: Evidence-based principles for the design of multimedia instruction. *American psychologist*, 63(8), 760.

Appendix A

<i>Participant Report of Learning Experience Means</i>					
<u>Mental Effort</u>	<u>Understanding</u>	<u>Learning</u>	<u>Concentration</u>	<u>Topic Interest</u>	<u>Engagement</u>
3.53 (1.34)	2.05 (1.18)	2.20 (1.23)	3.90 (1.38)	0.60 (0.03)	2.69 (1.32)

Note: Standard deviations are in parentheses.

Six video feedback questions were asked to better understand participants' impressions of the lesson they watched. Table reports rating means (with standard deviations in parentheses) of the following statements on a scale from scale from 1(not at all the case) to 5 (completely the case):

1. The video required very low or high mental effort.
2. The video was very easy or difficult to understand.
3. The video was very easy or hard to learn from.
4. The video required very little or much concentration.
5. The video made them more or less interested in the topic.
6. The video was very engaging or very boring.