

Adaptive Lesson Chunking Mitigates Cognitive Load and Improves Retention

Shaoheng Ren* and Abdullah Siddique

Pioneer Academy, USA

*Corresponding author: Shaoheng Ren, Pioneer Academy, USA

ARTICLE INFO

Received:  August 18, 2025

Published:  August 28, 2025

Citation: Shaoheng Ren and Abdullah Siddique. Adaptive Lesson Chunking Mitigates Cognitive Load and Improves Retention. Biomed J Sci & Tech Res 63(1)-2025. BJSTR.MS.ID.009839.

ABSTRACT

The present study investigates the value of adaptiveness in learning interfaces and has found results suggesting that optimizing, or 'chunking,' lessons in accordance with the learner's attentional capacity leads to better learning outcomes. A novel sort of experiment was conducted. Sixty adults from the United States aged 18-35 participated in the experiment and were randomly assigned among two groups. The experiment was conducted through a webpage (online), and the results also recorded therefrom. After being randomly assigned, all participants completed a preassessment, in order to gauge their cognitive capabilities. A classic 2-back, a psychomotor-vigilance test, and a slide-reading attention probe were used to operationalize cognitive capacity. If participants were in the experimental group, these metrics would be used as the determinant for lesson size; their lesson sizes ranged from 7-15 words, depending on attention scores. In contrast, the control group all received homogenous 10-word blocks. Study time was identical among both groups, and learners were instructed on the Toki Pona language through the interface. After instruction, immediate recall on a 10-item quiz was markedly greater for adaptive learners ($78\% \pm 15$) than for controls ($63\% \pm 20$), a large effect ($t(58)=3.2$, $p < .005$, $d \approx 0.8$) and with lesser variation. Since the preassessment indicated that both groups performed similarly on the cognitive metrics, the increase in the adaptive learners may be attributable to reduced extraneous load that came about as a result of personalized chunking. The present study utilizes one novel method of operationalization for cognition, which may hold significance. The results of the present study demonstrate that even relatively minor implementations of attention-based adaptation improve learning outcomes. The present study holds these implications for the designers of instructional interfaces and is contextualized in Cognitive-Load theory. A pathway to more efficient digital instruction is one of the chief significant implications of the present study.

Introduction

In the modern context, as mediums of instruction have grown increasingly digital, the possibility of incorporating new elements of learning, which were previously infeasible in a physical context, has come to light. Of the many novel possibilities introduced by a digital instruction interface, the present study fixates on adaptiveness in particular [1-3]. More specifically, the value thereof is examined through an experiment and the findings hold implications salient for instructional designers. It is thought that chunking lessons in sizes which accommodate the cognitive faculties of the learner will lead to best learning outcomes [1-3]. The basis for this assertion relies on previous findings, particularly those of cognitive-load theory [1]. The working memory, among other components implicated in learning, is characterized by such theories as a finite resource [4]. Pursuant to cognitive-load theory, the working memory allocates its resources to

three sorts of cognitive loads: intrinsic loads (relies upon the implicit difficulty of the material; whereas arithmetic may have low intrinsic load, calculus may rate higher), germane loads (relevant to integrating the information at hand into memory, critical for learning), and extraneous loads (relevant to dealing with distractions or taxing burdens, which could include poor instructional design). The mechanism proposed by the present study, and which is consistent with the findings of the experiment conducted, holds that by incorporating elements of adaptiveness into learning interfaces, such as by chunking lessons, the extraneous load implicit in an insensitive design is virtually minimized, allowing for a greater proportion of cognitive resources to be allocated to the germane load, which, in turn, improves learning outcomes. When instructional designs are insensitive to the learner's cognitive capacity, this may impose an extraneous load that detracts from germane loads, which, in turn, ultimately inhibits learning [5,6] (cf. the expertise-reversal effect [7]).

Methodology

The experiment conducted under the present study was executed through an online webpage and hosted sixty adults from the United States aged 18-35. Due to the anonymity implicit in conducting an experiment through an online medium, more specific demographics could not be recorded, and thus, more definitive descriptors of the sample could not be produced. Those demographics have critical consequences for generalizability and results of the experiment, which may be a limitation of the present study. Participants were adult volunteers who were incentivized to participate for a small-sum compensation; this compensation, though, was small and unlikely to have been a confounding factor in results. Among the sixty participants, there was equal and random assignment to either the control group (n=30) or the experimental one (n=30). The treatment for both groups was largely the same, and included an exhaustive preassessment of cognitive capacity, the instructional period, and a 10-item quiz to gauge the ultimate learning outcome. The only difference between the treatments, though, was merely that the instructional period for the experimental group was adapted, or chunked, with respect to the metrics derived from the preassessment. The control group all received uniform 10-word lessons, regardless of their performance, whereas the experimental group could fare anywhere from 7 to 15 words per chunked lesson. The preassessment battery was exhaustive and examined various elements of the learners' cognitive capacities, but it was deliberately not so intensive as to fatigue the learners, which could have had confounding consequences for performance. It's critical to emphasize, though, that although the sizes of the lessons varied between the treatments, the net volume of information supplemented was identical in that both groups were ultimately exposed to the same number of words.

The preassessment battery conducted three trials examining various components of cognition which may be implicated in learning. The metrics derived therefrom were weighted and scaled to a refined score used as the determinant for chunking lessons in the instructional period for the experimental group. Regardless of being in the experimental or in the control group, all learners completed these trials. The first trial of the preassessment battery was a classic n-back test of working memory with n = 2. A sequence of 30 uppercase letters was presented one-by-one each for 1½ seconds. Learners had to input their spacebar whenever the shown letter matched the one exactly two positions earlier. The first two letters were seeds with no response expected; thereafter, a 35% proportion of letters were deliberately programmed to be matches (the letter was identical to that two positions earlier) while the rest were non-matches. This sequence was designed to ensure that non-matches never accidentally matched the 2-back letter (preventing unintended matches). A brief tutorial was supplemented to ensure understanding, and then testing would commence. Metrics included number of correct detections ("hits"), missed detections ("misses," when a match occurred but the

learner failed to input it), and false attempts (inputs on non-matches). Working-memory performance was quantified with an overall score:

$$Accuracy_{2-back} = \frac{Hits}{Hits + Misses + False\ Alarms}$$

The second trial of the preassessment battery was a psychomotor vigilance test used to operationalize sustained attention and alertness that included 20 rounds. In each round, learners waited a randomized 2-7 seconds waiting for a grey circle to turn green, at which point, immediate input is expected. Two metrics were observed: reaction time (RT) in milliseconds and counts of false inputs. Any RT which exceeded 500 milliseconds was logged as a lapse in attention (a standard threshold for PVTs indicating momentary failure of vigilance). Vigilance was quantified with an overall score where N is the total trials (20) and L being lapses in attention, $w = 0.7$ and was implemented for weighing purposes:

$$Vigilance\ Score = \frac{1}{RT} - w \cdot \frac{L}{N}$$

The preassessment battery concluded with a novel metric for attention that operationalized for sustained attention and reading diligence. A slideshow was presented with 29 brief slides on cognitive topics and learners were informed to review all slides at their own pace. The time spent on each slide was logged, with a 1-second minimum to preclude accidental clicks. After all slides were reviewed, the interface collapsed these timings into a slide-attention score as per the following formula:

$$Score = 100(0.7S + 0.3[1 - CV]_+)P,$$

The wisdom was that if learners spent progressively less time on each slide, it was an indication of limited cognitive resources whereas consistency was used as a marker for broader resources. S is the sustained-focus fraction and is $d/29$ where d is the first slide where time spent falls below 60% of the baseline (mean of time spent on first three slides) or is simply equal to 29 if no such drop occurs. $CV = \sigma/\mu$ and is the coefficient of variation in time spent (lower CV indicates steadier reading, $[\cdot]_+ = \max(0, \cdot)$ prevents negative consistency scores). P is a spam-click penalty based on the proportion q of slides skimmed in less than 1 second: $P=1$ if $q=0$; $P=0.5$ if $0 < q \leq 0.30$; $P=0.2$ if $q > 0.30$. By this formula, diligent readers who maintain consistency, with minimal skipping, tended to score higher than those who were inconsistent, had erratic pacing, or were skipping tremendously. Of the three metrics, this was the most influential in the final weighed score and is, largely, a novel method of operationalizing for a component believed to be implicated in learning. Performance on these metrics converged by virtue of a singular formula into a difficulty index:

$$d = 0.70A + 0.20W + 0.10V$$

Where A is the attention on the slideshow, W is the 2-back accuracy, and V, the final PVT score. The formula gives preference to the slideshow metric, as it is the one believed to be most implicated in the process of learning. Of course, the other metrics are weighed in, in the event that one exhibits exemplary working memory or psychomotor capabilities or if severe deficiencies existed therein that may eclipse the performance on the slideshow. Lesson size L was set by the following linear rule, making the smallest chunked lesson 7 words and the largest containing 15:

$$L = \text{clip}_7^{15}(\text{round}[7 + 0.08d])$$

Though all learners completed the preassessment battery, the performance was only considered for those in the experimental group. This was done to ensure that any fatigue imposed by the battery would be controlled by having it administered to both groups. Immediately after the preassessment battery concluded, learners were admitted to the instructional period, where they were provided with a Quizlet-style learning interface that promoted learning through the use of flashcards. The flashcards contained words from the Toki Pona language; the present study selected this language since it was unlikely that learners would be familiar with it prior to the instructional period as they might have been had a more popular language been employed (for example, French).

The instructional period was relatively simple and included lessons with 10 second breaks enforced in between. Learners were untimed and instruction was self-paced; those within the control group received fixed 10-word blocks throughout the duration of the instructional period regardless of preassessment performance, whereas those within the experimental group had their lessons chunked according to those metrics, with anywhere from 7 to 15 words per lesson. The assumption follows that by tailoring and chunking lesson sizes to accommodate the assessed cognitive capacity of the learner,

that that learner would be liberated of the extraneous load implicit to an otherwise insensitive design, which, in turn, allows for greater incorporation of material into memory. Data were recorded concerning time spent on lessons to ensure engagement.

Immediately after the instructional period, a final assessment, a 10-item quiz, was administered to all learners. The quiz contained 10 multiple-choice questions randomly drawn from the Toki Pona words covered within the instructional period. Each question presented one Toki Pona word (written either in Toki Pona script or transliteration) as the prompt and with four English words as options (one key and three distractors). This format tests for retention of material taught within the instructional phase, and ultimate learning outcome and retention was operationalized by the present study by this axis. Once the participant selected an answer, the interface locked in the response and provided instant feedback, which may have been a confounding factor in performance in later questions. Each learner's accuracy was measured as the primary learning outcome.

Findings

The findings are largely consistent with the mechanism proposed by the present study to some extent. Figure 1 as imaged below represents the control group in blue and the experimental (adaptive, chunked) in green. The box-plots show (from left to right) the distributions of each group's performance on the 2-back task, the PVT, the slideshow metric, the accuracy on the 10-item quiz, and the average number of words per lesson supplemented throughout the instructional period (MW). (The control has none because they received only fixed 10-word blocks). Figure 2 as imaged below represents the same five metrics as shown in Figure 1 in the format of a table. Values suggest of approximate parity in baseline cognitive traits and a modest retention advantage for the experimental (adaptive, chunked) group (Figure 2).

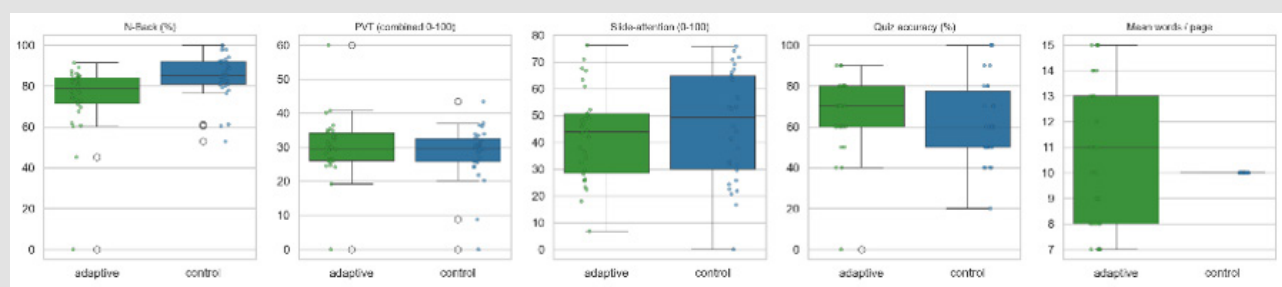


Figure 1.

| | Control | Adaptive |
|-----------------|---------|----------|
| nBack | 84.5 | 74.4 |
| PVT | 47.3 | 50.1 |
| Slide | 46.2 | 42.5 |
| Mean words/page | 10.0 | 10.8 |
| Quiz % correct | 63.0 | 67.7 |

Figure 2.

The preassessment metrics for cognitive capacity, as demonstrated by both Figures 1 & 2, hold similarly for both groups, which allows for the possibility of comparisons to be drawn with respect to the final learning outcome. The data demonstrates that both groups fared similarly in the preassessment battery; this common baseline allows not only for comparison, but for speculation on the disparity on the final learning outcome that advantaged the experimental (adaptive, chunked) group. A two-sample t-test on 2-back accuracy yielded no statistically significant difference ($p > 0.1$), confirming similar working memory. PVT followed suit, scores did not differ significantly. Median PVT scores hovered around the 40s for both groups with no statistical difference ($p > 0.5$). The sustained attention span yielded by the slideshow did not differ much either ($p > 0.2$). In summary, variance or minor numerical differences in these preassessment metrics is best attributed to noise (randomness), potential technical errors, or simply disengagement. There were some outliers excluded. Overwhelmingly, though, learners from both groups had similar cognitive profiles and is critical because it allows any disparity in performance to be attributed to the chunking rather than preexisting cognitive disparities between the groups.

Vastly, though, some statistical analyses revealed that the disparities in ultimate retention between the two treatments was statistically significant, while others deemed it as merely approaching significance. This introduced nuance, and invites further research on the subject matter to provide more conclusive results. Though the findings are, largely, consistent with the mechanism proposed by the present study, the extent to which is somewhat ambiguous and requires more thorough investigation. Though, it seems by large, that those with chunked, adaptive lessons enjoyed modestly greater learning outcomes than their counterparts who use insensitive interfaces; this likely occurs through the mechanism aforementioned by the present study. This does indeed suggest that adaptive interfaces, such as those using chunked lessons, may hold greater retention outcomes for learners as a whole.

Implications, Limitations and Discussion

Though the present study yields promising results with auspicious implications for adaptiveness and chunking in learning interfaces, further research is required to support the assertion at hand, that by introducing elements of adaptiveness, cognitive resources are liberated from extraneous loads and focused on germane loads, which, in turn, facilitates learning outcomes [1-3,8,9]. The final learning outcomes, are viewed statistically significant in one lens, but are approaching significance in another; thus, more research is required on this subject matter and the present study is best viewed as merely illuminating the potential of adaptiveness and chunking in learning interfaces [8,9]. One particular limitation concerning the present study is that the specific demographics of the sample of sixty adults, due to the anonymity presented by an online medium, could not be recorded adequately. It was merely known that they resided within the United States and ranged from 18-35. If the specific composition of the sample is unknown, this obviously presents an issue with generalizability.

For example, the highest level of education completed by the sample, which obviously is critical to a study of this nature, was left unrecorded. This is a limitation which must be considered. Similarly, though the present study has shown chunking to be efficacious, it is unclear if it fares equally in effectiveness. In conclusion, the present study holds interesting implications for adaptiveness and chunking in the context of learning, but this must be considered against the limitations herein listed or implied; further, more conclusive research is required to provide more definitive evidence, though the current as produced by the present study seems to suggest that adaptiveness and chunking in learning interfaces promote overall learning outcome. We invite further research, and look forward to all the potential that elements of adaptation in chunking may hold for learning [10].

References

1. Sweller J (1988) Cognitive load during problem solving: Effects on learning. *Cognitive Science* 12(2): 257-285.
2. Kalyuga S, Ayres P, Chandler P, Sweller J (2003) The expertise reversal effect. *Educational Psychologist* 38(1): 23-31.
3. Paas F, Renkl A, Sweller J (2003) Cognitive load theory and instructional design: Recent developments. *Educational Psychologist* 38(1): 1-4.
4. Sweller J, van Merriënboer JIG, Paas FGWC (1998) Cognitive architecture and instructional design. *Educational Psychology Review* 10(3): 251-296.
5. Sweller J, Ayres P, Kalyuga S (2011) *Cognitive load theory*. Springer.
6. Sweller J (2010) Cognitive load theory: Recent theoretical advances. In: JL Plass, R Moreno, R Brünken (Eds.), *Cognitive load theory*, Cambridge University Press, p. 29-47.
7. Kalyuga S (2007) Expertise reversal effect and its implications for learner-tailored instruction. *Educational Psychology Review* 19(4): 509-539.
8. van Merriënboer JIG, Sweller J (2005) Cognitive load theory and complex learning: Recent developments and future directions. *Educational Psychology Review* 17(2): 147-177.
9. De Jong T (2010) Cognitive load theory, educational research, and instructional design: Some food for thought. *Instructional Science* 38(2): 105-134.
10. Mayer RE, Pilegard C (2014) Principles for managing essential processing in multimedia learning: Segmenting, pre-training, and modality principles. In: RE Mayer (Edt.), *The Cambridge handbook of multimedia learning* (2nd Edn.), Cambridge University Press, pp. 316-344.

ISSN: 2574-1241

DOI: 10.26717/BJSTR.2025.63.009839

Shaoheng Ren. Biomed J Sci & Tech Res



This work is licensed under Creative Commons Attribution 4.0 License

Submission Link: <https://biomedres.us/submit-manuscript.php>



Assets of Publishing with us

- Global archiving of articles
- Immediate, unrestricted online access
- Rigorous Peer Review Process
- Authors Retain Copyrights
- Unique DOI for all articles

<https://biomedres.us/>