Pacing, Pixels, and Paper: Flexibility in Learning Words from Flashcards

Kara Sage, Joseph Rausch, Abigail Quirk and Lauren Halladay
Hamilton College, Clinton, NY, USA

ksage@collegeofidaho.edu; **joe.rausch@verizon.net**; **aquirk@characterlab.org**; **lehalladay@rochester.rr.com**

Abstract

The present study focused on how self-control over pace might help learners successfully extract information from digital learning aids. Past research has indicated that too much control over pace can be overwhelming, but too little control over pace can be ineffective. Within the popular self-testing domain of flashcards, we sought to elucidate the optimal level of user control for digital learning and compare learning outcomes between paper and digital flashcards. College students learned vocabulary from paper flashcards or one of several digital flashcard versions and were scored on their memory recall and asked about their perceptions of the learning process. With digital flashcards, students were randomly assigned to an automatic slideshow of cards with no user control, automatic slideshow with pre-set pauses, automatic slideshow where users could press the spacebar to pause at any time, or a self-paced slideshow with complete user control. Users reported feeling more in control when indeed having some control, but ultimately memory recall, cognitive load, and satisfaction were similar across the five versions. However, memory recall was positively related to user satisfaction with their specific flashcard set, and negatively related to users' perceived mental effort and difficulty. Notably, whether paper or digital, students showed individual variability in how they advanced through the words. This research adds to the educational literature by suggesting that paper and digital flashcards are equally viable options for students. Given differences between individual users and the connection between satisfaction and recall, individualistic options that offer, but do not force, some control over pace seem ideal. Paper flashcards may already include such options, and e-flashcards should offer similar adaptive features to appeal to a wide variety of users.

Keywords: computer, pace, self-control, learning, flashcards.

Introduction

As the current generation of digital natives enters college and the workforce, more and more of these students are turning to their computers, tablets, and smartphones to practice skills and gain knowledge. The increasing availability of digital materials, and online coursework more generally, has moved knowledge acquisition into a more autonomous realm than ever before. When students flip open their screen, they have a plethora of information at their fingertips. But, a challenge exists in that stu-

dents have to pace their own learning and individually deduce what information is important to extract. Online classes and learning materials often espouse this individualized nature as a key benefit; however, this also means that a certain level of individual motivation and self-regulation is necessary in order to successfully learn the material. Whether large-scale classes or small-scale learning aids, features of these digital offerings that encourage a successful, independent, and self-directed route to learning merit our examination.

Research has generally indicated that interactive features, referring to any feature that elicits user involvement, are beneficial to students' learning in the digital realm. In the present study, we evaluate one particular digital feature that may be beneficial in this regard - namely, the level of control over pace that a student has over the material. What is too little or too much? How does the level of control contribute to learning outcomes? And, how do students' perceptions of the learning platform relate to those outcomes?

We explore the answer to these questions within the context of a common learning aid for college students - flashcards to learn vocabulary. Flashcards have been a popular self-testing strategy for students for many years (Kornell & Bjork, 2008) and have continuously evolved; e-books and online studying materials commonly now offer digital flashcards (e.g., Griggs, 2012 and Myers, 2009, as cited in Golding, Wasarhaley, & Fletcher, 2012). Though popular with students, there is a dearth of research investigating both paper and digital flashcards. We seek to examine what level of control over pace is optimal for digital flashcards as well as elucidate how student learning compares between digital and traditional paper flashcards. Given the current shift in student learning to the online, individualized environment, clarifying how much control a student should have over the pace of a digital study aid will provide an important practical application for future program design.

Control as Interactivity in Digital Learning

In recent years, researchers have become increasingly focused on what aspects of the digital environment might lead to successful knowledge acquisition, with interactivity and control over pace/content emerging as potential candidates (e.g., Burdick & Willis, 2011; Davidson & Goldberg, 2009; Greenhow, Robelia, & Hughes, 2009). These concepts often go hand-in-hand; when a program is interactive, it inherently involves some level of control over the material as a result of the requisite user involvement. This user involvement often increases a computer program's utility as a learning tool for students. For instance, Zhang, Zhou, Briggs and Nunamaker (2006) discovered that e-learning with an interactive video (allowing for control over pace and content access) resulted in superior learning outcomes when compared to e-learning with a non-interactive video, no video, or even a traditional classroom environment. The interactive video allowed users to tailor the experience to their own pace as well as engage more actively with the content. Students were also most satisfied with the interactive video. However, students were learning about Internet search engines in this experiment; given that the content was digital itself, perhaps learning it digitally was most relevant.

Research into digital textbooks has also supported that active control over material positively affects learner attitude (Yoon, Kim, & Choi, 2014). Likewise, interactive online assessments result in more complex reasoning when compared to static or less active online assessments (DeBoer et al., 2014). Despite generally positive support, there are some mixed opinions on the value of interactivity (e.g., Bezjian-Avery, Calder, & Iacobucci, 1998; Ghose & Dou, 1998). Shackel (1991) implicated three aspects of interactivity that determine a program's success. Students must think that the program is successfully teaching the material (utility). They must find it enjoyable (likeability) and not be overburdened by its functionality (usability). Furthermore, adaptive features should allow for self-guided discovery of the material. Inherent in these adaptive features is some level of user control, aiming for a low-demand/high-utility fusion.

Motivation may also play a role here. Controlling the flow of information, such as by pausing or replaying sections, may enhance intrinsic motivation (Domagk, Schwartz, & Plass, 2010; Wang & Reeves, 2007; Zhang, 2005). For instance, when using YouTube videos to learn a skill (Jaffar, 2012; Lee & Lehto, 2013), the user can select videos and view them in various sections. Similarly, recent research highlighted that segmented videos, with natural pauses to think about the material, improve learning and are less cognitively demanding for college students when compared to continuous videos (Cheng, Huang, Shadiev, Hsu, & Chu, 2014). Digital learning platforms with control options respond more organically to the user's needs; this responsivity may not exist in a traditional classroom or when no user interface is available.

Controlling a slideshow

Students frequently learn from slideshows, the predominant form of lecture in a traditional college classroom. Slideshows seem reminiscent of flashcards, given that one can similarly flip back and forth between slides. When viewing slides on one's own computer, a student likely has control over the speed of presentation. Given the research on digital features, an important consideration is how much the user (versus the program) should be in control of the slideshow presentation. Research has offered conflicting results (Lawless & Brown, 1997; Scheiter & Gerjets, 2007). While some research suggests that self-pacing options can help learners acquire knowledge from slideshows (e.g., Mayer & Chandler, 2001), other research has suggested that fully self-paced learning does not result in superior learning outcomes. For instance, Sage and Baldwin (2014) reported that students learned about novel actions with higher fidelity after viewing slides automatically advancing by the computer at a pre-determined pace (in this case, the pace of another user) when compared to self-paced slides. Follow-up work (Sage, 2014) confirmed that memory recall was superior following a slideshow program reliably paced by the computer when compared to self-controlled slideshows.

One possibility is that full self-control may cognitively overload the learner. Cognitive overload refers to instances in which there are no cognitive resources left over for completing the task at-hand (Vandewaetere & Clarebout, 2013). In Sage and Baldwin (2014) and Sage (2014), learners viewed over 500 slides to learn just a few actions and had to click the mouse each time to advance to a new slide. Attention was undoubtedly split between the images on screen and the mouse in-hand, and participants may have simply become fatigued or frustrated by the pace, perhaps leading to this cognitive overload. Scheiter and Gerjets (2007) have confirmed that usability issues can interfere with one's learning. Vandewaetere and Clarebout (2013) also suggested that, given the requirements of interactive programs, self-control over the flow of information can result in extraneous cognitive load. Thus, self-control or self-pacing is not inherently always helpful; its utility depends on the form of control and likely the specific content. For instance, Hoffler and Schwartz (2011) found that learners benefitted from self-pacing when the learning material consisted of animations but not when static images.

To more closely investigate what sort of user control is optimal for students' learning from slideshows, Sage, Bonacorsi, Izzo, and Quirk (2015) investigated how college students learned about new actions (magic tricks) from four slideshow types. Results of this study demonstrated the importance of self-control in the form of pause capability. Memory recall was higher in the slideshow versions with pause capability when compared to the version with no user control. Additionally, most students reported that the version with no user interaction was their least favorite while the version with the most user control (where slides advanced automatically but users could click the mouse to pause at any point) was their favorite format to learn from. The influence of mindset was also evident; these positive reports occurred in spite of only ~25% of users actually using the interactive click-to-pause feature of the slideshow; in other words, 75% of users in this click-to-pause condition just viewed a computer-controlled version and never engaged the paus-

ing feature. Thus, the mindset of being in control, not the self-pacing specifically, seemed to increase users' liking of the program. This finding seems to support the idea that an option for self-pacing is important and appreciated by the learner, but that such a feature should not be unduly forced upon the learner.

Taking these various findings into consideration, a blend of system and user control may be optimal for a digital flashcard program, as it allows some room for freedom while also generally guiding someone through a program. Hannafin (1984) also supported this blend, describing how, if left entirely up to the learner, the user might take an unintended, albeit individualized, path through the information that may not result in successful learning. Furthermore, research has shown that user attitudes affect learning outcomes (e.g., Kettanurak, Ramamurthy, & Haseman, 2001); likewise, Sage et al. (2015) implicated that students want and like control over pace even if they do not use the feature. Thus, it seems likely that there is an optimal level of control over pace that mirrors a Goldilocks effect - requiring neither too much nor too little, but providing that "just right" level of control over the learning process.

As mentioned previously, the current study will focus on this issue of control over pace within the context of flashcards. We now turn to some relevant literature on flashcards, to provide background information on what we already know about how digital and paper flashcards compare, and what features have already been deemed successful or not in this realm.

Flashcards: What Do We Know?

Background

Though research on flashcards is relatively scarce, flashcards have been shown to increase users' verbal abilities, such as with vocabulary and reading (e.g., Kupzuk, Daly, & Anderson, 2011). In an examination of 415 Introduction to Psychology students (Golding et al., 2012), students using flashcards on all exams had significantly higher exam scores than students not using or only sometimes using flashcards. Overall, ~70% of students used flashcards for at least one exam, with the vast majority using traditional paper flashcards.

Most often, college students use flashcards to self-test, with the purpose of determining how well they have learned the information. Interestingly, Hartwig and Dunlosky (2012) found that, though endorsement of self-testing was significantly related to college GPA (self-testing seemed to be positive metacognitive skill), the use of flashcards was not. Other research has pointed to students' motivation for self-testing being focused on a desire to evaluate progress more than to improve (Kornell & Son, 2009; Wissman, Rawson, & Pyc, 2012). It is also possible that flashcards were not used as effectively as they could have been by students. Flashcards also promote surface-level memorization (Davis, 2013); to develop a solid grasp of course material, students must instead think deeply about its intricacies. However, memorization can be useful. For instance, Tan and Nicholson (1997) reported that the emphasis on rapid word recognition encouraged by flashcards was beneficial for 7-10 year old below-average readers.

Features

Self-control has been investigated to some extent already in relation to flashcards, however, primarily only with paper flashcards and also defined in different ways than just control over pace. Kornell and Bjork (2008) were interested in how control, in the form of deciding whether or not to remove cards, would affect student learning. Participants were given two 20-item sets of English-Swahili translations on paper cards and had the option either to remove cards or not while initially going through the words. Participants performed more poorly on memory tests when they had dropped cards from the stack. Follow-up studies ascertained that this behavior might have

resulted from students removing cards too quickly, specifically after getting the translation correct only once. This work suggests that some program control over digital flashcards might be helpful, as some words were easily forgotten when the option to remove a word was left entirely up to the user. Interestingly, some digital flashcard programs already have a feature built in to discard a card quickly after a designated number of responses. Given this prior research on paper flashcards, perhaps digital programs should have a more stringent criterion so as to not remove cards too early to the detriment of the learner.

Besides self-control, another feature to consider is type of presentation. Kornell (2009) had undergraduate students study GRE word-synonym pairs on paper cards. When comparing different stacking techniques (one stack versus four smaller stacks), participants remembered more from the large stack but also falsely believed that the smaller stacks would produce better results. This finding suggests that some guidance about the use of flashcards might be helpful for students, as they mistakenly thought that one version would be superior. This research again highlights that some program control over flashcards might be optimal, as students may not always self-select the option that results in the best learning outcomes.

Digital versions

A small body of work has focused specifically on digital flashcards. In older research, Fuson and Brinko (1985) tested children's ability to learn math facts from microcomputers and flashcards. Children in 2nd-4th grade learned problems via either paper flashcards or the computer. Groups learned their math facts using their assigned method for four weeks, and then all children switched methods. Math performance increased for all children, but there was not a significant difference in outcomes between groups. It is notable that performance was hitting a plateau at week 4, but all groups again showed an increase in knowledge after switching to the new type of studying. Children also reported liking the computer more because they felt like they could not cheat as easily and believed it was faster and easier to use. Other older research looking into comparisons of computer drills and traditional paper flashcards showed little difference in test performance as well (Klein & Salisbury, 1987). This research suggests that digital and paper flashcards might be similar in their learning outcomes for students, though this research is quite outdated for today's digital standards. It can also be noted that students liked the computer more, and other past research has shown that learner perceptions and outcomes are related (e.g., Kettanurak et al., 2001).

In more recent research, Forbes et al. (2013) explored how a particular list of words affected learning. Third and fourth graders were given either a list of unknown words or a mixed list of known and unknown words to learn from an online flashcard program. Interestingly, the interspersing of unknown into known words resulted in slower learning. Thus, how you set up an online flashcard program is of importance; certain formats may be more successful than others.

Similarly, Davis (2013) investigated the effectiveness of digital flashcards for exam preparation. Students enrolled in a course on construction materials were given iPod Touches with the Flashcards Deluxe App. Beyond flashcards, the app included options to create multiple-choice questions and spelling quizzes. Nearly all students reported that the flashcards had helped. Furthermore, the majority of students not using the app believed that performance on past exams would have increased with flashcard use. Students also thought the digital flashcards were easy to carry around and more convenient than notes. When comparing two classes using flashcards to eight prior classes not using flashcards, there were mostly no differences in exam scores. In fact, a few exams from one semester with flashcards showed decreases in scores, though the author reported that this particular class might have had unusually poor self-regulation. Davis concluded that flashcards did not help or hinder learning and that flashcards were likely a better studying tool

than no studying tool at all. It seems possible that the introduction of the iPod Touch itself might have created a novelty effect that reduced potential for learning.

Summary

When considering this research on flashcards, it seems that students may not know the most effective way of using flashcards to bolster learning. Given this conclusion, a blend of user and program control for a digital flashcard program might be optimal in helping guide students down an effective path of learning. Also, it seems currently unclear how paper and computer-based flashcards compare in their usefulness, particularly in our digital age.

Current Study

The present research examined what level of control over the pace of digital flashcards is optimal for learning outcomes while also focusing on the utility of paper versus digital flashcards. Knowledge acquisition is quickly becoming more of an online, individualized endeavor, and clarifying the features to incorporate into such programs is of great importance.

To address these aims, we employed five flashcard versions. Version 1 was paper, and versions 2-5 were digital with varying levels of control over the pace of the presentation. Past research has supported that full user control can, at times, lead to cognitive overload or usability issues (e.g., Sage & Baldwin, 2014; Scheiter & Gerjets, 2007) but also that control over pace is superior to no interactivity (e.g., Zhang et al., 2006). To make a thorough comparison, we developed programs at different points along this full user to full program control spectrum. At one end, version 2 was the digital, self-paced equivalent of paper flashcards (full user control over pace). At the other end of the spectrum, version 3 was a completely automated digital flashcard presentation (full program control over pace). Versions 4 and 5 struck blends between user and program control by allowing for pause capability (in light of some of the past research on pauses, e.g., Sage et al., 2015). Version 4 automatically paused after a subset of words and required a press of the spacebar to continue (blending user and program control). Version 5 could be paused at any time via the spacebar (full user control if the person clicked the spacebar repeatedly, full program control if the user did not touch the spacebar at all, or some blend).

To determine the success of these various flashcards, we measured outcomes in two ways: one, how well students remembered the vocabulary, and two, students' perceptions of their flashcard experience. Students were asked two questions tapping cognitive load (perceived difficulty and mental effort), along with two questions on overall satisfaction and perceived control over the flashcards.

Hypotheses: Paper versus digital

Hypothesis #1: Using the self-paced paper and digital versions will result in similar memory recall and perceptions of the flashcard experience. These versions provide the same user-controlled presentation, and, given this sample's age and familiarity with digital media, this outcome seemed sensible. The minimal past research comparing the two forms is also in line with this hypothesis (Fuson & Brinko, 1985; Klein & Salisbury, 1987).

Hypothesis #2: Using paper flashcards will produce better memory recall and satisfaction than using digital flashcards (all versions combined). We believed this outcome would be the case, given prior research showing the helpfulness of paper flashcards (e.g., Golding et al., 2012) and that too much or too little control is linked to poorer learning outcomes (e.g., Sage et al., 2015; Zhang et al., 2006). It also seemed possible that the novelty of digital flashcards (e.g., Davis, 2013) would increase participants' ratings of difficulty/mental effort, though we expected per-

ceived control to be consistent given that some of the options were user-controlled similarly to the paper flashcards.

Hypotheses: Control over pace

Hypothesis #3: Using self-paced digital flashcards will lead to better memory recall and satisfaction than using the other digital versions (three versions combined). In line with Shackel (1991), this self-paced program was designed with utility, likeability, and usability in mind. We attempted to reduce the cognitive overload potential from other research (e.g., Sage et al., 2015) by having a minimal number of button clicks, and thus presumably keeping difficulty/mental effort low. Thus, though fully self-controlled, we believed this outcome would occur given that the self-paced version could most readily be tailored to the user and mirrored the typical functionality of paper flashcards. Given that the other versions incorporated varying degrees of computer control, this overall outcome seemed likely.

Hypothesis #4: Using automatic paced digital flashcards will lead to lower memory recall and be less satisfying than using the two versions with pausing functionality. This hypothesis stems from research supporting the benefits of presenting segmented information over continuous presentations (e.g., Mayer & Chandler, 2001), noting that some control over pace is helpful. Though the three programs are likely similar in difficulty, it seemed possible that mental effort would be higher in the automatic pace version given the lack of a pause during which to review the words, and that perceived control would be higher when having some control via pauses.

Hypothesis #5: Using the press-to-pause version will result in superior memory recall and satisfaction when compared to using the automatic pausing every five words. The spacebar press allows for the pauses to be tailored to the user (as well as the amount of program versus user control over pace), and thus we expected higher perceived control ratings in the press-to-pause version. Even without any pauses, prior work has shown that the mindset of being in control is beneficial to the learning process (Sage et al., 2015). Research on flashcards specifically has also shown that one large stack (akin to press-to-pause) results in superior learning to smaller stacks (akin to the pausing every five words), even in spite of user conceptions to the contrary (Kornell, 2009). Given similar functionality, we expected difficulty/mental effort to be equivalent.

Hypothesis: Performance-perception relations

Hypothesis #6: Memory recall and user perceptions of their flashcard experience will relate. Specifically, we hypothesized that satisfaction and perceived control would positively relate to memory recall, while mental effort and difficulty would negatively relate to memory recall. This hypothesis is in line with past research linking attitudes and performance (Kettanurak et al., 2001; Teo, Oh, Liu, & Wei, 2003).

Methods

Participants

Participants were 110 undergraduate students (median age = 19 years, 74% female, 75% White) at a Northeastern liberal arts college, receiving extra credit for psychology courses. The first 30 participants used paper flashcards. The remaining 80 participants were randomly assigned to one of four digital versions (n = 20 per condition): self-pace, automatic pace, automatic pace/pause every five words, press-to-pause. Three other individuals participated, but were excluded due to not following rules (n = 1) and experimental error (n = 2).

Materials

Words

We selected words from GRE (Graduate Record Examination) practice lists. Our reason for utilizing this source was to create a list of challenging words for an 18-22 year old undergraduate liberal arts audience. Four undergraduate researchers reviewed hundreds of words from practice GRE lists. They removed words that all (or almost all) knew and selected a subset of nouns that were easily translatable into one-word synonyms. Forty-two words remained; twenty-seven words were randomly selected for use (2 practice words, 25 test words) and the remaining 15 words were not used. Microsoft Excel's random number generator determined word order. All participants received words in this order (see Table 1).

Table 1

Words and Synonyms

Words	Synonyms
Sycophant (practice word)	Groupie
Leviathan (practice word)	Colossal
Effrontery	Nerve
Largess	Donation
Probity	Integrity
Equipoise	Balance
Interstice	Opening
Acumen	Guile
Dearth	Absence
Maelstrom	Chaos
Calumny	Lie
Peccadillo	Indiscretion
Misanthrope	Cynic
Tyro	Beginner
Chicanery	Dishonesty
Imprecation	Curse
Encomium	Compliment
Caprice	Fickleness
Precipice	Cliff
Quiescence	Calm
Torpor	Apathy
Magnanimity	Generosity
Diatribe	Tirade
Vim	Activity
Turpitude	Depravity
Iconoclast	Critic
Depredation	Plunder

Flashcards

Flashcards had the vocabulary word on one side (with phonetic pronunciation), and synonym definition on the other side. We opted for this design as it had minimal encoding requirements (a student must only remember one word and not a more complex definition) and was similar to the set-up on vocabulary practice lists (Kornell, 2009). To help distinguish which side was the word versus synonym, we printed words in crimson and synonyms in orange font. Words appeared on 4 x 6 white flashcards (see Figure 1).
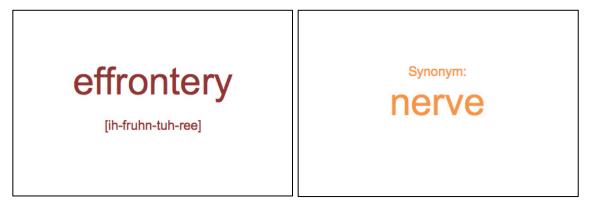

Figure 1. Example of a paper flashcard

Computer program

A computer programmer wrote a desktop Java application to display the digital flashcards on a 21.5-inch iMac desktop computer. An initial instruction screen prompted the researcher to enter participant number and flashcard version. Upon hitting enter, the flashcard program launched (see Figure 2). For all versions, the computer program recorded the card duration (how long the card stayed on the screen before the participant/program moved on) and the number of pauses and user-instigated flips, if any.

Figure 2. Example of a digital flashcard

In the self-paced version, participants used the mouse to click the "flip card" button. When ready to move on to a new word, they clicked "next card". Pace was entirely in the user's control. In the automatic pace version, participants had no control over pace. The program displayed each card for 30 seconds, with five flips of the card (five seconds with the word on the screen, five seconds with the synonym on the screen, etc.). We wanted this automatic version's timing to be similar to

the original paper version, thus we coded participants' use of the paper flashcards (from videotape). On average, participants flipped the card 4.50 times (SD = 2.39), which we rounded to five flips. Approximate duration on a single side before flipping the card was 5.20 seconds, which we rounded to five seconds. In the automatic pace/pause every five words version, the timing matched the automatic pace version, with pauses inserted after the 5th, 10th, 15th, and 20th cards. We opted to insert this pause after every five words, as five is the traditional bottom boundary of items remembered in working memory (7+/- 2; Miller, 1956). Users pressed the spacebar when ready to continue. In the press-to-pause version, the timing was again matched, but the user could pause the program at any time by pressing the spacebar. As in the self-paced version, the user could flip the card any number of times, and was instructed to click "resume slides" when they were ready to resume automatic pacing.

Measures

Memory recall

Memory tests #1 and #3 consisted of a sheet of paper with two columns of 25 blank lines. The participant was instructed to recall all words and corresponding synonyms in any order. Each participant received a correct pairs score. The only difference between memory tests #1 and #3 was timing - memory test #1 occurred immediately after the learning phase; memory test #3 occurred later after a distraction task, as a measure of retention. Memory test #2 was a sheet of paper with all vocabulary words listed in column 1 and 25 blank lines in column 2. The participant was told to recall all synonyms. Each participant received a correct synonyms score. Memory test #2 was meant to be a bit simpler than memory tests #1 and #3, as well as more similar to everyday tests a student might receive in class.

Learner perceptions of flashcard experience

Via a written survey, participants rated four questions, on a 7-point scale, with 1 as extremely not (e.g., extremely dissatisfied) and 7 as extremely so (e.g., extremely satisfied). Two questions tapped cognitive load: How much effort did it cost you to complete this flashcard task? How difficult was it for you to learn these words from flashcards? Two questions tapped other perceptions: How satisfied are you with this flashcard format? How in control of the flashcards did you feel?

Self-report questions for cognitive load and satisfaction are typical of learning research, and have been shown to be reliable (e.g., Paas, Tuovinen, Tabbers, & van Gerven, 2003; Teo et al., 2003). Similar questions on mental effort, difficulty, and satisfaction, as well as 7-point scales, are common (e.g., Kablan & Erden, 2008; Moreno & Valdez, 2005; Vandewaetere & Clarebout, 2013). The distinction between germane (question on difficulty) and intrinsic (question on mental effort) cognitive load is also important in tapping different types of load. In addition, we included a question on perceived control, as that was a key aspect to our study.

For participants experiencing the digital flashcard versions, we also asked open-ended questions about what they liked/disliked about the program. This qualitative data supplemented our quantitative measures, aiding us in understanding differences, or lack thereof, between the digital flashcard versions generally and also the levels of control specifically.

Additional survey questions

Participants completed a written survey asking for demographic and past exposure information, e.g., familiarity with the 25 words (M = 3.36, SD = 3.63). Participants also reported their Scholastic Aptitude Test (SAT) verbal score, if remembered, and completed a verbal ability test, where

they provided definitions for 15 words that often appear on SAT practice lists. The same team of undergraduate researchers selected these words (see Table 2), and scored them out of 15, indicating the number of correct definitions. The purpose of collecting this background information was to provide relevant sample information and to allow for analyses to control for aspects like verbal ability. However, analyses did not vary as a function of these variables, and will thus not be further discussed in the results.

Table 2

Words Testing for Verbal Ability

SAT Words
Counterfeit
Remuneration
Tangent
Abasement
Blasphemy
Nuance
Absolution
Plethora
Plumage
Accolade
Lampoon
Plaudit
Harbinger
Enigma
Laceration

Procedure

Participants completed the consent process and then were instructed that they would be viewing a set of flashcards. After being given basic instructions about the functionality of their version and told "Working within these few rules, please use the flashcards in whatever way seems natural to you to learn the words," participants reviewed two practice words and then told the experimenter when they were ready to continue. At this point, participants were either directed to the stack of paper flashcards or to the computer screen to view their 25 words.

For all conditions, participants viewed one word at a time and could move only forward in words (i.e., no back-tracking). For the traditional paper flashcards, participants dropped a card in a box once they were finished with it. For the digital versions, the computer advanced or the participant clicked on "next card", depending on the version.

Once finished with the flashcards, participants completed the other measures in order: memory test #1, memory test #2, distraction task (10-minute Sudoku puzzle), memory test #3, perceptions survey, qualitative questions (digital versions only), and background survey. Participants were then debriefed and thanked.

Analytical Strategy

To ascertain differences between flashcard versions, we opted for two between-subjects multivariate analyses of variance (MANOVA). The MANOVA strategy was selected given its inclusion of multiple dependent variables in a single analysis. For each MANOVA, the seven dependent measures were correct pairs from tests #1 and #3, correct synonyms from test #2, and self-reported satisfaction, mental effort, control, and difficulty.

Paper versus digital

The first MANOVA focused on paper flashcards versus its self-paced digital counterpart. This test centered on hypothesis #1 - that participants would learn equally well from these two versions. The second MANOVA looked across all five versions and employed Helmert contrasts. Contrast 1 compared paper flashcards to all digital flashcards, speaking to hypothesis #2 (paper flashcards as superior).

Control over pace

Contrast 2 compared the self-paced digital version to the three automated versions, providing information on hypothesis #3 (self-paced as superior). Contrast 3 compared the automatic pace version to the two versions with pause functionality, speaking to hypothesis #4 (pause functionality as superior). Finally, contrast 4 compared the two versions with pause functionality, providing information on hypothesis #5 (press-to-pause as superior).

Supplementary data

We reviewed supplementary data to shed light on the support, or lack thereof, for hypotheses #1-5. Specifically, we examined card duration and flipping behaviors of participants by condition, practice trials, and participants' open-ended responses about what they liked or disliked about the digital programs.

Performance-perception relations

To speak to hypothesis #6, we conducted a series of correlations between the aforementioned seven dependent measures to determine if performance (three memory scores) and perception (four survey questions) were meaningfully related.

Results

Hypotheses #1-5: Do Learning Outcomes and Perceptions Differ Across Versions?

For the first analysis, the omnibus MANOVA was not significant, Wilks' Lambda = 0.81, $F(7,42) = 1.38$, $p = .24$. Looking at the univariate tests, only perceived control was significant, $F(1,48) = 4.70$, $p = .04$. Participants felt more in control of the self-paced digital version than the paper version. Thus, hypothesis #1 (no difference) was primarily supported given that participants did not differ on six of the seven measures.

For the second analysis, speaking across all five flashcard versions, the omnibus MANOVA was marginally significant (Wilks' Lambda = 0.68, $F(28,358.37) = 1.46$, $p = .065$. Looking at the univariate tests, perceived control was significant, $F(4,105) = 5.53$, $p < .001$. The univariate test on mental effort was marginally significant, $F(4,105) = 2.21$, $p = .07$. For contrast 1 (paper v. digital), no test reached significance, thus primarily disagreeing with hypothesis #2; only perceived control was initially hypothesized as equivalent.

For contrast 2 (self v. auto pacing), only the contrast on perceived control reached significance, p = .001. Learners felt more in control of the self-paced digital version than the automatic versions. Findings for hypothesis #3 were thus mixed; we expected perceived control to be higher and mental effort/difficulty to be consistent, however we also expected memory recall and satisfaction to be higher for self-paced learners. For contrast 3 (fully automatic v. pause capability), no tests were significant, thus discordant with hypothesis #4; only perceived difficulty was expected to be similar. Lastly, for contrast 4 (pause every five words v. press-to-pause), participants rated perceived control as higher, p = .001, and mental effort as higher, p = .02, in the press-to-pause version than the version pausing every five words, thus providing mixed findings related to hypothesis #5; we expected control to be higher and difficulty to be similar, however we also expected memory recall and satisfaction to be higher along with mental effort being similar. See Table 3 for descriptive statistics.

Table 3

Descriptive Statistics: Dependent Variables

	Test #1: Correct Pairs	Test #2: Correct Synonyms	Test #3: Correct Pairs	Self-Reported Satisfaction	Self-Reported Effort	Self-Reported Control	Self-Reported Difficulty
	M (SD)	*M (SD)*	*M (SD)*	*M (SD)*	*M (SD)*	*M (SD)*	*M (SD)*
Paper	4.67 (3.54)	11.27 (4.88)	5.63 (3.71)	3.87 (1.48)	4.83 (1.21)	3.40 (1.67)	5.10 (1.16)
Self Pace	3.55 (3.90)	8.90 (6.21)	4.65 (4.88)	4.30 (1.56)	4.20 (1.85)	4.55 (2.06)	4.95 (1.70)
Auto Pace	4.05 (2.58)	9.55 (5.24)	5.35 (4.04)	4.25 (1.29)	4.75 (1.21)	2.95 (1.28)	5.10 (1.37)
Pause 5 Words	4.90 (4.36)	11.10 (5.20)	5.80 (4.20)	3.80 (1.47)	3.90 (1.52)	2.40 (1.05)	4.90 (1.41)
Press-to-Pause	4.20 (3.68)	9.35 (5.67)	5.00 (4.21)	3.95 (1.36)	4.95 (1.19)	4.10 (1.92)	5.20 (1.36)

Note: Memory test scores out of 25; Self-report ratings out of 7.

To look at our hypotheses from a different angle, we considered if recall differed by version based on the serial position of the words - words at the start and end of a list are often best remembered (Murdock, 1962). We conducted another MANOVA with version as the fixed factor, again with Helmert contrasts. For memory tests #1 and #3, we scored first words (words 1-7), middle words (words 8-18), and last words (words 19-25). For all memory tests, we scored synonyms in the same way. Looking to the contrasts, for contrast 1, on memory test #1, participants remembered more synonyms from the last seven words when using paper over digital cards, p = .05. This result fits nicely with Hypothesis #2 (paper flashcards as superior). Contrast 2 was also significant, again for memory test #1. Self-pacing resulted in fewer recent words, p = .04, and synonyms, p = .05, remembered, when compared to auto-pacing, in contrast to Hypothesis #3 (self-paced as superior).

Hypotheses #1-5: Supplementary Data on Flashcard Versions

Flips and duration

To clarify why differences may or may not have emerged between flashcard versions, we analyzed the number of card flips and the duration that each card was viewed by version. To mimic our earlier analysis, we conducted a MANOVA with version as the fixed factor, Helmert contrasts, and the two dependent variables of average card duration and number of flips. The omnibus MANOVA was significant, Wilks' Lambda = 0.66, F(8,208) = 6.04, p < .001. The univariate test on flips per card was not significant, F(4,105) = 1.03, p = .40, but the univariate test on card duration was significant, F(4, 105) = 7.46, p < .001. For card duration, contrast 2 (self v. auto pacing) was the only test to reach significance, p < .001, suggesting that participants spent less

time per card in the self-paced version than in the other digital versions. This finding could relate to the lack of support for hypothesis #3 (self-pacing as superior); people may have advanced through the words too fast to reap a benefit of self-control over pace. See Table 4.

For the digital versions, the computer program generated an extra layer of information - the duration spent specifically on the word versus synonym. The videos from the paper version did not allow for this more nuanced coding, given grainy imagery and participants' tendency to block our view (e.g., by putting cards on their laps). Thus, for just the digital versions, we re-ran the MANOVA with the three dependent variables of word duration, synonym duration, and number of flips. The omnibus test was significant, Wilks' Lambda = 0.32, $F(9,180.24) = 12.11$, $p < .001$. Only the univariate test for synonym duration was significant, $F(3,76) = 29.63$, $p < .001$. When looking at the contrasts, for contrast one, word duration, $p = .03$, and synonym duration, $p < .001$, were significant, such that individuals spent less time on both card types in the self-paced versions than in the other versions. This finding again may relate to the lack of support for hypothesis #3 (self-pacing as superior). For contrast two (fully automatic v. pause capability), no tests were significant, perhaps highlighting that similar use contributed to the lack of support for hypothesis #4 (pause functionality as superior). For contrast three (pause every five words v. press-to-pause), synonym duration was significant, $p < .001$, such that individuals spent less time on the synonyms in the press-to-pause version than in the version that paused every five words. This result informs hypothesis #5 (press-to-pause version as superior), supporting one reason why perceptions of control and mental effort were higher in the press-to-pause version.

Table 4

Descriptive Statistics: Card Flips and Duration

	Flips	Card Duration (sec)	Word Duration (sec)	Synonym Duration (sec)
	M (SD)	*M (SD)*	*M (SD)*	*M (SD)*
Paper	4.49 (1.44)	23.38 (12.21)	-	-
Self Pace	5.75 (4.32)	18.24 (11.89)	12.46 (7.85)	5.78 (4.72)
Auto Pace	5 (0)	30.01 (0.0007)	15.001 (0.0004)	15.004 (0.0004)
Pause 5 Words	5 (0)	32.35 (3.84)	15.003 (0.0004)	17.35 (3.84)
Press-to-Pause	5.22 (2.25)	28.08 (10.09)	15.73 (6.19)	12.35 (5.53)

Duration - pause every 5 words

How long did individuals actually pause when the program halted every 5 words? We were curious if individuals were just clicking instantly, or if they were stopping to potentially contemplate the words. With the exception of one outlying individual that paused approximately 17 seconds each time, individuals were very quickly hitting the spacebar to continue, averaging 1.58 seconds (SD = 1.78). Participants did not seem to use the pre-programmed pause to mull over the words; this finding may explain the lack of support for hypothesis #4 (pause functionality as superior).

Press-to-pause use

One other factor of interest was whether or not participants actually used the press-to-pause mechanism. Inaction would indicate that viewers experienced the same presentation as automatic pace viewers, even if mindset varied. When looking at the 20 participants within the press-to-pause condition, nine participants (45%) never paused, thus engaging full program control. Three individuals (15%) paused a single card while two individuals (10%) used the pausing feature on 4-5 cards, thus experiencing a blend of user and program control. The remaining six individuals (30%) paused all cards, taking full user control and using the press-to-pause feature to its full capacity. Participants' open-ended responses also reflected that some individuals took advantage of

this option (e.g., "Always [paused] so that I had as much time as I needed to picture the words.") while others did not (e.g, "I never paused the slideshow because I felt like seeing each slide three times was enough"). This information sheds light on the absence of support for hypothesis #5 (press-to-pause version as superior), as a majority of participants never or rarely enacted the press-to-pause feature.

Also, given that these numbers were weighted on the two extremes (not using the press-to-pause feature at all versus using it on every card), we did a supplementary MANOVA comparing memory recall and perceptions between the nine "never used it" participants and the six "always used it" participants. The omnibus test was not significant, Wilks' Lambda = 0.47, $F(7,7) = 1.14$, $p = .43$. When looking at the univariate tests, only the test on perceived control was significant, $F(1,13) = 5.54$, $p = .035$. Perhaps unsurprisingly, participants using the control feature felt more in control ($M = 5.17$, $SD = 2.14$) than participants not using the control feature ($M = 3.11$, $SD = 1.27$).

Was practice helpful?

Logically, individuals who used the practice flashcards to explore the pacing feature (glancing more than once at each side) might perform better than participants not using the practice cards well. For the three digital versions where the user had some control, most individuals (75-85%) did practice, but learning outcomes did not differ as a function of practice. (This analysis was a MANOVA, with practice group as the fixed factor and memory scores as the dependent variables.) Comparatively, with paper flashcards, only 39% of individuals practiced, perhaps reflecting that students were more familiar with paper cards. When comparing individuals practicing versus not practicing on paper, there was a significant difference on memory test #1 score, $F(1,21) = 7.08$, $p = .02$. The 14 participants practicing briefly remembered more correct pairs ($M = 6.67$, $SD = 4.72$) than the nine participants practicing extensively ($M = 2.92$, $SD = 1.94$). Perhaps some participants were already comfortable with the paper format, rendering practice trials unnecessary. Ultimately, given that differences were limited to one test, practicing was not consistently indicative of a learning advantage or disadvantage.

Qualitative data - what did participants like or dislike?

Common remarks from participants' written open-ended responses appear in Table 5. Generally, users liked the flexibility of pausing and disliked the lack of pause capability when not offered. When pauses were offered, users implicated general features of the program as the negative aspects. When not offered, users seemed to comment on other general features as the positive aspects.

Table 5

Positive and Negative Remarks

	Common Positive Remarks	Example Statements	Common Negative Remarks	Example Statements
Self Pace	Ability to flip/pause; simplicity	"ability to flip as many times as possible" "pretty simple, easy to use"	General features; desire for physical cards	"Couldn't go back to old words" "I think I'm a kinesthetic learner so I probably would have done better with physical flashcards"
Auto Pace	General features	"it was functional" "I liked the use of synonyms"	Lack of a pause; desire more control	"didn't like how I couldn't go at my own pace" "At times I did want to pause"
Pause 5 Words	General features	"I liked that the words and synonyms were two different colors" "I liked that it repeated the word 3 times"	Dislike pause; desire more control	"I didn't like being able to pause in between words" "I do not like that the program cannot be self-paced and we cannot pause it."
Press-to-Pause	Appreciate flexibility/ choice to pause; simplicity	"It was user friendly as it allowed me to take as much time as I needed per slide" "I liked how clear everything was" "easy, simple and intuitive"	General features	"That you couldn't go back to look at past cards."

Hypothesis #6: How do Performance and Perception Relate?

To determine interrelations between performance and perception, we ran correlations between the seven dependent measures across our full sample. Firstly, it can be noted that the memory measures were strongly positively intercorrelated (all $r(108) > 0.77$, all $p < .001$). Given this consistency, we only present the performance-perception correlations for memory test #2, as memory test #2 most closely mirrored a test students would take in their daily lives, thus providing the most meaningful take-away message.

Findings largely supported hypothesis #6. Satisfaction with the medium was positively correlated with number of correct synonyms, $r(108) = 0.30$, $p = .002$. Likewise, how much mental effort was expelled was negatively correlated with number of correct synonyms, $r(108) = -0.22$, $p = .02$, as was how difficult a participant perceived the task to be, $r(108) = -0.58$, $p < .001$. Interestingly, perceived control of the flashcards was not significantly correlated with any learning outcome (p's > .48), though it can be noted that r-values were in a positive direction.

There was also coherence in perceptions. Satisfaction was negatively correlated with difficulty, $r(108) = -0.41$, $p < .001$, and marginally with mental effort, $r(108) = -0.17$, $p = .08$. Difficulty and mental effort (our two measures of cognitive load) were positively correlated, $r(108) = 0.32$, $p = .001$. Interestingly, perceived control was also positively related to satisfaction, $r(108) = 0.38$, $p < .001$, even though perceived control was not related to mental effort or difficulty.

See Table 6 for a summary of outcomes.

Table 6

Summary of Outcomes of Hypotheses

	Primarily Supported?	Primarily Negated?
Hypothesis #1: Using the self-paced paper and digital versions will result in similar memory recall and perceptions of the flashcard experience.	x	
Hypothesis #2: Using paper flashcards will produce better memory recall and satisfaction than using digital flashcards (all versions combined).		x
Hypothesis #3: Using self-paced digital flashcards will lead to better memory recall and satisfaction than using the other digital versions (three versions combined).		x
Hypothesis #4: Using automatic paced digital flashcards will lead to lower memory recall and be less satisfying than using the two versions with pausing functionality.		x
Hypothesis #5: Using the press-to-pause version will result in superior memory recall and satisfaction when compared to using the automatic pausing every five words.		x
Hypothesis #6: Memory recall and user perceptions of their flashcard experience will relate.	x	

Discussion

This research investigated how one particular feature of digital programs - amount of control over pace - influenced learning outcomes and perceptions. Given the scarcity of research on flashcards, we also aimed to highlight any difference in college students' learning of vocabulary from traditional paper versus digital flashcards. When comparing different levels of control over pace, and paper and digital flashcards more broadly, perceptions and learning outcomes were strikingly similar. However, perceptions of satisfaction, mental effort, and difficulty were related to learning outcomes in the full sample, suggesting that individual perceptions are directly linked to performance. Evidence thus supported that the individual user should decide what type of flashcard and pace control is best for him or her. This finding helps us better understand the utility of pacing features - flexibility appears key, but there may not be one "optimal" type. As for new forms of flashcards, digital versions seem quite promising and are no more or less difficult to use than paper flashcards.

Paper cards are the traditional option, already accepted by many users. They also appeal to our sense of simplicity - easy to create and use. It is important to note that we are not seeking to negate their use in any way. After all, learning outcomes were equivalent across our flashcard versions. We are instead claiming that digital options appear to be just as viable, and, regardless of flashcard type, some flexibility in how we use flashcards (e.g., in pace, options to flip a card, etc.) seems optimal given that individual users have different preferences. Though paper cards are the "tried-and-true" method and offer some inherently flexible features, digital options are increasing in prevalence and may become a more preferred method of study for today's "digital native" student population. Digital cards offer their own unique set of benefits, such as being more accessible (perhaps on one's phone) or tied directly to an e-book where a student can save vocabulary words and make e-flashcards with one mouse click as they read. They also present unique challenges - e.g., media multitasking detracts from attention (Ralph, Thomson, Cheyne, & Smilek, 2014). Furthermore, there is an established way that students tend to approach the creation and

use of paper cards in their studying. The same cannot yet be said about digital cards. Scrolling through different flashcard apps on the Internet or on the iPad quickly informs us that there are many options available. Research like ours provides more information regarding what form of digital flashcard might be most ideal and user friendly. If some students are going to be turning to their screens for flashcards, determining the features of successful programs to maintain attention and enhance learning outcomes is critical. Though there is much to still be explored in the digital realm, and paper flashcards remain dominant, e-flashcards hold promise and merit our examination.

Reflecting on Hypotheses

The equivalence of learning outcomes across flashcard versions did not align with our hypotheses. We believed that, when the self-paced characteristic was held constant, paper and digital cards would be equivalent, and this notion was confirmed. However, we believed that paper cards would be preferred over digital cards in general, given their history of use, and that controlling the pace would enhance learning outcomes when compared to some digital versions having little or no control. One potential reason for this null finding is that all participants were learning the same words from the same cards, and perhaps platform and pacing simply did not differentially affect their learning. Further, vocabulary learning is an activity that college students engage in regularly, and they might be used to learning these words in flexible ways. It would not be surprising if students were already learning words both on paper (textbooks, flashcards, notes) and digitally (on phones, tablets, computers). Additionally, our particular sample was high in verbal ability. Average self-reported verbal SAT score of the 66 students reporting their score was 683 (out of 800, SD = 65.92). At this time, the College Board's website reported a score of 680 as representing the 94th percentile. Perhaps, given this pre-existing strength, participants were primed to learn words successfully across any condition.

In addition, our supplementary flashcard usage data suggested that students approached each flashcard set similarly. For instance, there was no difference in the number of flips of a card across versions; users saw the words and the synonyms a similar number of times. Furthermore, even when pauses were built-in to encourage the user to reflect on the words, students instead opted to immediately advance forward. Similarly, about half of participants in the press-to-pause version never pressed the spacebar. In addition, the basic display functionality likely instantiated segments across all versions (Mayer & Chandler, 2001; Schaffer & Hannafin, 1986). Essentially, no matter the functionality or level of control, students navigated the task in a similar way.

We did note an interesting finding relating to the recency effect of short-term memory (Murdock, 1962). Students remembered more recent words with the paper flashcards than with the digital flashcards. As a whole, people may be more comfortable with paper flashcards. Moreover, the difference could be due to the context of word recall. In the paper version, participants used paper flashcards and took a memory test on paper, thus providing similar contexts of learning and reproducing the words. Further, students using the self-paced digital cards showed less of a recency effect when compared to students using the automatic-paced cards. As participants neared the end of the flashcards in the self-paced version, perhaps they proceeded more quickly or their attention waned from the constant clicking. Conversely, participants in the automatic-paced versions viewed the cards for a consistently longer period, as controlled by the computer. Notably, these differences were only observed in conjunction with memory test #1. With the passing of time, the recency effect likely diminished.

In addition, we noticed some differential patterns in the versions in which participants had more control versus less control over their own learning process. These patterns ultimately pointed to students varying in their preferences; perhaps what was "just right" for one person was different from what was "just right" for another person. In our sample, a subset of individuals pressed the

spacebar to gain control of the program when given the option. Those users tended to adjust their viewing habits relative to other conditions as well, such that they spent comparatively less time on the synonyms. Similarly, our digital self-paced users also averaged less time per card than the other digital versions, suggesting that some users might like to move more quickly through the cards. In addition, when looking at the open-ended responses from our participants in the fully computer-controlled version, just as many reported that it was too fast as reported it was too slow. Logically, the best speed of presentation not only varies by user but by other factors such as pre-existing knowledge, difficulty level, and importance of the task. These influences vary in weight by person, resulting in variable optimal speeds of presentation. The wide range in use in the self-paced version showcases this finding. The average flips per card varied from 2.40 to 16.24 (M = 5.75, SD = 4.32) and the average duration per card varied from 5.54 to 50.07 seconds (M = 18.24, SD = 11.89). Programs should include flexible levels of control over pace so as to better match individual desires. This flexibility is inherently a feature of paper cards, and digital programs should mirror this option.

There were also notable differences in perceived control across the flashcard versions. Most of these findings were not surprising. Users felt they had more control over self-pacing than auto-pacing. Also, when comparing the automatic versions that either paused every five cards or via the press of the spacebar, participants naturally felt more in control of the latter. Interestingly, people felt more in control of the digital self-paced version when compared to the paper flash-cards. It seems possible that, as current college students, many of our participants regularly inter-act with screen technology, perhaps even more than they interact with print, and thus feel more in control of it. Also, the plethora of portable digital platforms currently available (e.g., tablets and cell phones) has made using the screen an on-the-go activity and increasingly convenient for us-ers. Furthermore, given their history, it is possible that students had more expectations about pa-per flashcards. Perhaps the relative novelty of digital cards led individuals to feel more in control, as they did not feel like they were missing out on some traditional flashcard aspect. Lastly, we can also consider the actions involved in using paper versus digital cards, as the click of a button arguably takes less effort than flipping a card.

It is also important to keep in mind that prior work (Sage et al., 2015) highlighted the notion that mindset matters - even if a user opts to not control the pace when possible, the mindset of being in control can aid learning outcomes. Though we did not replicate that finding (our press-to-pause version was not viewed as superior), we can consider the relevant application. If we incorporate simple self-pacing features as an option into programs, it is possible that users will benefit. We saw great variability in our self-paced users' navigation of our program, and our users had vary-ing perceptions of perceived control. Under different circumstances (e.g., with a more complex task), these pacing features might become increasingly important.

In addition, it is worthy to note that our self-paced users did not seem to be burdened; they were not placed under higher cognitive load than other users because of the demands of self-control. Self-reported difficulty and mental effort were consistent across versions, with the exception of considering when to press the spacebar as requiring somewhat more mental effort than just click-ing to advance after five slides. The self-paced version being equal in difficulty/mental effort to the more computer-controlled versions contrasts with past research; others have pointed to com-puter control being superior to self-control in slideshows, perhaps resulting from extraneous cog-nitive load or usability issues (e.g., Sage & Baldwin, 2014). Given that these aspects can deter learning (Scheiter & Gerjets, 2007), it was imperative to create a self-paced option that did not create additional challenges for users. In our program, buttons were clearly labeled, the function-ality was simple, and the presentation was brief. Thus, full user control did not seem to signifi-cantly inhibit learning, though it can be noted that some program control outside of pace still ex-isted; the program determined the order of the words.

Returning to Shackel's (1991) ideas, students did not seem overtaxed by our program's functionality (usability). Difficulty and mental effort were reported around the neutral mid-point of the scale across all versions. Students did not dislike the program (likeability); however, ratings did hover near the neutral mid-point and thus the program was also not well liked. Based on the open-ended responses, there were consistent reasons across versions that might have led to this neutral perception, such as the inability to return to prior words (e.g., "I cannot scroll back so no opportunity to review"). In terms of utility, the intent of the program was to help students remember words. We asked students how many words they had known definitions for before the study, and they reported having prior knowledge of 3.36 words (SD = 3.63). Given that our means for test #2 (where students supplied definitions) were much higher than that, it appears that the program was indeed helpful in learning the words (utility).

Finally, it is important to emphasize the relationships we found between performance and users' perceptions of their flashcard experience. When collapsed across versions, students' satisfaction was positively related to their memory recall, while perceived effort and difficulty were inversely related to memory recall. Thus, being satisfied with a learning strategy and perceiving the program as easy to use might have contributed to one's ability to successfully learn. These results are in line with past research showing that attitude relates to learning outcomes (Kettanurak et al., 2001; Teo et al., 2003). Interestingly, though perceived control positively related to satisfaction, it did not significantly relate to learning outcomes, though correlations were in a positive direction. Thus, while control might be one aspect in determining satisfaction, it might carry less importance in determining outcomes than mental effort and difficulty. If a program can be simple and satisfying with no control, it might be viewed just as positively as a program that is simple and satisfying with control. Ultimately, for individuals seeking control, that option should be available. Again, the material might matter. For learning vocabulary, a familiar task to college students, perhaps control is less related to performance. For a more complex, unfamiliar skill (e.g., building a bookcase), it is possible that self-control would increase in importance.

Practical Implications: Recommendations for Educational Materials

Flashcard programs, and educational materials in general, should offer adaptive features that allow students to tailor their learning experience. Since learning outcomes were directly tied to students' own perceptions of satisfaction, difficulty, and mental effort, this flexible learner-centered framework is key. While paper cards are inherently individualized, digital flashcards currently offer a wide range of different experiences for users. Determining the nature of this digital experience is critical for future program design to ensure that digital flashcards have just as much an impact on education and student learning as traditional paper flashcards.

Given our research, we believe that self-control features must be designed with simplicity and flexibility in mind, with the goal of avoiding cognitive overload. A main flaw of past digital designs (e.g., Sage, 2014; Sage & Baldwin, 2014) may have been that self-control was too effortful, creating usability issues (Scheiter & Gerjets, 2007). In our work, self-paced learners were just as satisfied and did not report their program as comparatively more difficult to use. Further, pause functionality is a recommended feature. This choice aligns with our design and past work showcasing that presenting information in segments is helpful for learning (e.g., Mayer & Chandler, 2001; Schaffer & Hannafin, 1986).

Overall, blending system and user control in digital programs seems optimal (Hannafin, 1984). Even in our self-paced version, the program guided the overall flow of information while the user could adjust pace. Future research should more fully investigate the ideal blend, but, in the meantime, programmers should consider this amalgam along with usability issues. A poor combination

could be frustrating and not meet Shackel's (1991) ideas of likeability, utility, and usability. A prime blend could reduce cognitive load and lead to positive perceptions/outcomes.

Lastly, it also seems critical that screen technology be incorporated into students' education. Students have been described as "tethered" to their technology (Turkle, 2011); phones are practically considered another limb. As such, learning is frequently becoming an online activity where independent navigation of the material is key. Students may already have an individual preference for paper or digital materials, or it may depend on the context (e.g., phone on the go, paper at home). Making their own choices about platform, and encouraging students to play an active role in determining their own speed of learning, can promote intrinsic motivation, and, in turn, their learning (Domagk et al., 2010; Wang & Reeves, 2007; Zhang, 2005).

Ultimately, as described, digital platforms should incorporate a learner-centered framework, which responds to the diversity of users (McCombs & Vakili, 2005). And, given that too much freedom might be frustrating and that students may not always self-select the options best for learning (Hannafin, 1984; Kornell, 2009; Vandewaetere & Clarebout, 2013), some underlying program structure should be paired with options for user control in order to assist students in successfully navigating an online, individualized learning environment.

Limitations and Future Directions

In our study, students were given one learning phase and tested immediately thereafter. This process does not mimic real life. For instance, a student learning vocabulary might study their flashcards several times before feeling prepared, and an exam might occur weeks later. The one-time exposure to our flashcard set may not have been adequate for full memorization. Given that students never remembered all words, we might conclude that they would have studied more if it were part of their coursework. Furthermore, to mimic typical college exams, our memory tests required pen and paper. However, given research on context-dependent memory (Godden & Baddeley, 1975), students may have remembered more words if tested digitally. Additionally, though successfully testing memory recall, the format of our tests might not have reflected typical exams. In support of this idea, some participants expressed surprise when handed the first memory test where they had to write in both the words and synonyms. In-class exams tend to give more context clues, more similar to our second test that asked for just synonyms. To discover differences by test type, it would be interesting to compare what flashcards might be most helpful for low versus high stakes tests. Given that paper flashcards are more established while digital flashcards are more novel, individuals might prefer to rely on the "tried-and-true" paper flashcards and/or appreciate more control when faced with a high stakes test.

In the future, it would be interesting to use digital and paper flashcards on a larger scale, such as paired with a particular textbook across a semester. Davis (2013) looked at digital flashcards across the semester, comparing success to past semesters without flashcards. But, there was no comparison to paper flashcards. An ideal study would match variables that could affect learning (e.g., same class, semester, base knowledge) and then offer the different flashcard versions to evaluate short-term (e.g., on a given test) and long-term (e.g., on a retention task the subsequent semester) effects. Such large-scale research might inform educators whether or not these flashcards would be helpful to include with e-books.

Furthermore, a different blend of user and program control could be optimal. Our various options allowed us to examine different levels of control over pace, but these are not the only options. It could be interesting to allow students to make their own cards; creating such cards could boost learning outcomes, particularly if in conjunction with instructor-created flashcards (Cihon, Sturtz, & Eshleman, 2012). Though, it can be noted that often e-textbooks come with pre-made flashcards. A fair number of participants also reported that they wished they could have gone back-

wards and returned to prior words. Further, we used a computer to display the digital flashcards, but most students today are using laptops or tablets or even cell phones. Thus, we could more closely look at these newer platforms to see if there is an optimal platform for displaying these flashcards. There are other features of technology that could potentially add to flashcard programs, such as quizzes, as used in Davis (2013).

Lastly, given our focus on the importance of the individual learner, it seems critical to look further at the interplay between program and user. When considering personality, Devaraj, Easley, and Crant (2008) found that individual traits play a role in perceived usefulness of technology; for instance, neuroticism was negatively associated with perceived usefulness while agreeableness was positively related to perceived usefulness. It seems possible, for instance, that individuals high in agreeableness might engage more with novel digital cards than individuals low in agreeableness. Furthermore, when considering control options, Homer and Plass (2014) reported that students high in executive function performed more successfully after self-guided presentations while students low in executive function performed more successfully after program-guided presentations. This result might also relate to the self-regulation differences that Davis (2013) mentioned. Similarly, Gay (1986) pointed out that learners with pre-existing knowledge on a topic benefitted more from self-control then learners with little knowledge. Given that our sample was high in verbal ability, it would also be interesting to look at other people with varying verbal abilities to see if they would find different control features more helpful. It is also possible that motivation or current anxiety could directly interact with what type of program is "just right" for an individual learner.

Conclusion

Educational materials are rapidly becoming available on digital devices, and students are often independently navigating their way through the material. Given the challenge of self-directed learning, design features must help set a student up for success and reduce the risk for usability issues and cognitive overload. We evaluated one particular feature that might be useful in helping students navigate a digital tool - namely, different levels of control over the pace of the presentation. The present work demonstrated that the specific level of control over pace might be less important than the perceptions of the individual user regarding the specific learning medium (e.g., satisfaction, difficulty). A blend of user and program control may be optimal for digital learning aids, but flexibility in features may be more important; some users may learn well from full user control while others may need a more guided experience. We also reported that traditional and digital platforms can be equally successful in helping students learn, at least in the domain of flashcard use. Future work must establish what program features are optimal for creating a successful digital learning environment for individual learners, as well as continue to compare digital to traditional learning materials to ascertain their feasibility in education.

Acknowledgement

References

Bezjian-Avery, A., Calder, B., & Iacobucci, D. (1998). New media interactive advertising vs. traditional advertising. *Journal of Advertising Research, 38*, 23-32.

Burdick, A., & Willis, H. (2011). Digital learning, digital scholarship, and design thinking. *Design Studies, 32*(6), 546-556. doi:10.1016/j.destud.2011.07.005

Cheng, P., Huang, Y., Shadiev, R., Hsu, C., & Chu, S. (2014). Investigating the effectiveness of video segmentation on decreasing learners' cognitive load in mobile learning. *New Horizons in Web Based Learning, 8699*, 122-129. doi:10.1007/978-3-319-13296-9_14

Cihon, T., Sturtz, A., & Eshleman, J. (2012). The effects of instructor-provided or student-created flash-cards with weekly, one-minute timings on unit quiz scores in introduction to behavior analysis courses. *European Journal of Behavior Analysis, 13*, 47-57. doi:10.1007/s10864-007-9038-x

Davidson, C., & Goldberg, D. (2009). *The future of learning institutions in a digital age*. MIT Press.

Davis, K. (2013). Using electronic flashcards for student self-evaluation of readiness for exams. *Proceedings of the 120th American Society for Engineering Education Annual Conference and Exposition*.

DeBoer, G., Quellmalz, E., Davenport, J., Timms, M., Herrmann-Abell, C., Buckley, B., Jordan, K., Huang, C., & Flanagan, J. (2014). Comparing three online testing modalities: Using static, active, and interactive online testing modalities to assess middle school students' understanding of fundamental ideas and use of inquiry skills related to eco-systems. *Journal of Research in Science Teaching, 51*(4), 523-554. doi: 10.1002/tea.21145

Devaraj, S., Easley, R., & Crant, J. (2008). How does personality matter? Relating the five-factor model to technology acceptance and use. *Information Systems Research, 19*(1), 93-105. doi:10.1287/isre.1070.0153

Domagk, S., Schwartz, R., & Plass, J. (2010). Defining interactivity in multimedia learning. *Computers in Human Behavior, 26*, 1024-1033.

Forbes, B., Skinner, C., Black, M., Yaw, J., Booher, J., & Delisle, J. (2013). Learning rates and known-to-unknown flash-card ratios: Comparing effectiveness while holding instructional time constant. *Journal of Applied Behavior Analysis, 46*(4), 832-837. doi:10.1002/jaba.74

Fuson, K., & Brinko, K. (1985). The comparative effectiveness of microcomputers and flash cards in the drill and practice of basic math facts. *Journal for Research in Mathematics Education, 16*(3), 225-232. doi:10.2307/748395

Gay, G. (1986). Interaction of learner control and prior understanding in computer-assisted instruction. *Journal of Educational Psychology, 78*(3), 225-227. doi:10.1037/0022-0663.78.3.225

Ghose, S., & Dou, W. (1998). Interactive functions and their impacts on the appeal of Internet presences sites. *Journal of Advertising Research, 38*, 39-43.

Godden, D., & Baddeley, A. (1975). Context-dependent memory in two natural environments: On and land and under water. *British Journal of Psychology, 66*, 325-331. doi:10.1111/j.2044-8295.1975.tb01468.x

Golding, J., Wasarhaley, N., & Fletcher, B. (2012). The use of flashcards in an introduction to psychology class. *Teaching of Psychology, 39*(3), 199-202. doi:10.1177/0098628312450436

Greenhow, C., Robelia, B., & Hughes, J. (2009). Learning, teaching, and scholarship in a digital age: Web 2.0 and classroom research: What path should we take now? *Educational Researcher, 38*(4), 246-259. doi:10.3102/0013189x09336671

Griggs, R. (2012). *Psychology: A concise introduction* (3rd ed.). Worth Publishers.

Hannafin, M. (1984). Guidelines for using locus of instructional control in the design of computer-assisted instruction. *Journal of Instructional Development, 7*(3), 6-10. doi:10.1007/bf02905753

Hartwig, M., & Dunlosky, J. (2012). Study strategies of college students: Are self-testing and scheduling related to achievement? *Psychonomic Bulletin and Review, 19*, 126-134. doi:10.3758/s13423-011-0181-y

Hoffler, T., & Schwartz, R. (2011). Effects of pacing and cognitive style across dynamic and non-dynamic representations. *Computers and Education, 57*(2), 1716-1726. doi:10.1016/j.compedu.2011.03.012

Homer, B., & Plass, J. (2014). Level of interactivity and executive functions as predictors of learning in computer-based chemistry simulations. *Computers in Human Behavior, 36*, 365-375. doi:10.1016/j.chb.2014.03.041

Jaffar, A. (2012). YouTube: An emerging toll in anatomy education. *Anatomical Sciences Education, 5*, 158-164. doi:10.1002/ase.1268

Kablan, Z., & Erden, M. (2008). Instructional efficiency of integrated and separated text with animated presentations in computer-based science instruction. *Computers and Education, 51*, 660-668. doi:10.1016/j.compedu.2007.07.002

Kettanurak, V., Ramamurthy, K., & Haseman, W. (2001). User attitude as a mediator of learning performance improvement in an interactive multimedia environment: An empirical investigation of the degree of interactivity and learning styles. *International Journal of Human-Computer Studies, 54*, 541-583. doi:10.1006/ijhc.2001.0457

Klein, J. D., & Salisbury, D. F. (1987). A comparison of a microcomputer progressive state drill and flashcards for learning paired associates. *The 1987 Proceedings of Selected Research Papers*, pp. 401-418. Association for Educational Communications and Technology.

Kornell, N. (2009). Optimising learning using flashcards: Spacing is more effective than cramming. *Applied Cognitive Psychology, 23*, 1297-1317. doi:10.1002/acp.1537

Kornell, N., & Bjork, R. (2008). Optimizing self-regulated study: The benefits and cots of dropping flashcards. *Memory, 16*, 125-136. doi:10.1080/09658210701763899

Kornell, N., & Son, L. (2009). Learners' choices and beliefs about self-testing. *Memory, 17*(5), 493-501. doi:10.1080/09658210902832915

Kupzuk, S., Daly, E., & Andersen, M. (2011). A comparison of two flash-card methods for improving sight-word reading. *Journal of Applied Behavior Analysis, 44*, 781-792. doi:10.1901/jaba.2011.44-781

Lawless, K., & Brown, S. (1997). Multimedia learning environments: Issues of learner control and navigation. *Instructional Science, 25*, 117-131.

Lee, D., & Lehto, M. (2013). User acceptance of YouTube for procedural learning: An extension of the technology acceptance model. *Computers and Education, 61*, 193-208. doi:10.1016/j.compedu.2012.10.001

Mayer, R., & Chandler, P. (2001). When learning is just a click away: Does simple user inter-action foster deeper understanding of multimedia messages? *Journal of Educational Psychology, 93*(2), 390-397. doi:10.1037/0022-0663.93.2.390

McCombs, B., & Vakili, D. (2005). A learner-centered framework for e-learning. *Teachers College Record, 107*(8), 1582-1600. doi:10.1111/j.1467-9620.2005.00534.x

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63*(2), 81-97. doi:10.1037/h0043158

Moreno, R., & Valdez, A. (2005). Cognitive load and learning effects of having students organize pictures and words in multimedia environments: The role of student interactivity and feedback. *Educational Technology Research and Development, 53*, 35-45. doi:10.1007/bf02504796

Murdock, B. J. (1962). The serial position effect of free recall. *Journal of Experimental Psychology, 64*(5), 482-488. doi:10.1037/h0045106

Myers, D. (2009). *Psychology* (9th ed.). Worth Publishers.

Paas, F., Tuovinen, J., Tabbers, H., & van Gerven, P. (2003). Cognitive load measurement as a means to advance cognitive load theory. *Educational Psychologist, 38*, 63-72. doi:10.1207/s15326985ep3801_8

Ralph, B., Thomson, D., Cheyne, J., & Smilek, D. (2014). Media multitasking and failures of attention in everyday life. *Psychological Research, 78*(5), 661-669. doi:10.1007/s00426-013-0523-7

Sage, K. (2014). What pace is best? Assessing adults' learning from slideshows and video. *Journal of Educational Multimedia and Hypermedia, 23*(1), 91-108.

Sage, K., & Baldwin, D. (2014). Looking to the hands: Where we dwell in complex manual sequences. *Visual Cognition, 22*(8), 1092-1104. doi:10.1080/13506285.2014.962123

Sage, K., Bonacorsi, N., Izzo, S., & Quirk, A. (2015). Controlling the slides: Does clicking help adults learn? *Computers and Education, 81*, 179-190. doi:10.1016/j.compedu.2014.10.007

Schaffer, L., & Hannafin, M. (1986). The effects of progressive interactivity of learning from interactive video. *Educational Communication and Technology, 34*(2), 89-96.

Scheiter, K., & Gerjets, P. (2007). Learner control in hypermedia environments. *Educational Psychology Review, 19*, 285-307. doi:10.1007/s10648-007-9046-3

Shackel, B. (1991). Usability - Context, framework, design, and evaluation. In B. Shackel & S. Richardson (Eds.), *Human factors for informatics usability* (pp. 21-38). Cambridge University Press, UK.

Tan, A., & Nicholson, T. (1997). Flashcards revisited: Training poor readers to read words faster improves their comprehension of text. *Journal of Educational Psychology, 89*(2), 276-288. doi:10.1037/0022-0663.89.2.276

Teo, H., Oh, L., Liu, C., & Wei, K. (2003). An empirical study of the effects of interactivity on web user attitude. *International Journal of Human-Computer Studies, 58*, 281-305. doi:10.1016/s1071-5819(03)00008-9

Turkle, S. (2011). *Alone together: Why we expect more from technology and less from each other*. New York, NY: Basic Books.

Vandewaetere, M., & Clarebout, G. (2013). Cognitive load of learner control: Extraneous or germane load? *Education Research International, 2013*, 1-11. doi:10.1155/2013/902809

Wang, S., & Reeves, T. (2007). The effects of a web-based learning environment on student motivation in a high school earth science course. *Educational Technology Research and Development, 55*(2), 169-192. doi:10.1007/s11423-006-9016-3

Wissman, K., Rawson, K., & Pyc, M. (2012). How and when do students use flashcards? *Memory, 20*(6), 568-579. doi:10.1080/09658211.2012.687052

Yoon, S., Kim, M., & Choi, J. (2014). Effects of digital textbook's interactivity on the learning attitude: With a focus on the tablet pc-based digital textbooks of social studies and science. *The Journal of the Korea Contents Association, 14*(2), 205-222.

Zhang, D. (2005). Interactive multimedia-based e-learning: A study of effectiveness. *The American Journal of Distance Education, 19*(3), 149-162. doi:10.5392/jkca.2014.14.02.205

Zhang, D., Zhou, L., Briggs, R., & Nunamaker, J. (2006). Instructional video in e-learning: Assessing the impact of interactive video on learning effectiveness. *Information & Management, 43*, 15-27. doi:10.1016/j.im.2005.01.004

Biographies

Kara D. Sage is an assistant professor of psychology at the College of Idaho. She completed this research while spending two years as a visiting assistant professor of psychology at Hamilton College. She earned her Ph.D. in psychology from the University of Oregon, along with a master's degree in education from the University of Washington. In addition to teaching a wide variety of classes in psychology, she actively conducts research on new technologies, their psychological correlates, and applications for education.

Joseph R. Rausch is a candidate for juris doctor at Columbia Law School. Before attending law school, he earned a B.A. in psychology and public policy from Hamilton College.

Abigail L. Quirk graduated this May from Hamilton College with a B.A. in psychology. She is currently working as a research coordinator in Angela Duckworth's educational psychology lab at the University of Pennsylvania. She plans to continue on to graduate school in a few years in either education or psychology, and hopes to one day enter the world of school administration.

Lauren Halladay graduated this May from Hamilton College, with a psychology major and education studies minor. She is originally from Rochester, NY, but will be attending the University of Oregon as a school psychology Ph.D. candidate this coming fall. She enjoys working with children and hopes to someday help students reach their highest potential.