Schedule of practice matters.

Does it matter for video-based software training?

Master Thesis

December 2021

Supervisor: Hans van der Meij

Second supervisor: Johannes Steinrücke

Author: Karlygash Nuketayeva

Faculty of Behavioural Management and Social Sciences

University of Twente

TABLE OF CONTENTS

Acknowledgments

Writing this thesis was an exciting journey, which I could not have completed without the generous support, patience, and encouragement of many people in my academic and personal circles during my time at the University of Twente.

First and foremost, I would like to express my deep gratitude to my research supervisor Dr. Hans van der Meij for giving me the opportunity to do this research, for challenging me, inspiring, and supporting me through this journey. Dr. van der Meij's counter-arguments and guidance helped me examine this fascinating topic more comprehensively and in greater depth. His questioning of my initial argumentation and results challenged me do more and further increased my curiosity at every stage of my research work. I thank Dr. van der Meij for believing in me from the time I first took his class at the University of Twente as a newcomer to educational science through working on this research under his supervision. It's been an absolute honour to be supervised by Dr. van der Meij, whose work is well-known and recognized by scientists around the world.

I am grateful to my second supervisor Dr. Johannes Steinrücke for carrying out the uneasy task of doing the final review of a completed thesis, where the thinking process is done. I truly appreciate Dr. Johannes Steinrücke for helping me make my thesis better by giving me insightful yet motivating advice to polish the writing and improve the presentation of my data analysis and results. I thank both of my supervisors for supporting me in my desire to retain all of my research findings, which I believed were valuable and important to keep.

I would like to express my gratitude to all academic and administrative staff at the University of Twente for not only enriching me academically but also demonstrating their understanding to all students completing their studies during the COVID-19 pandemic. I am also thankful to my relatives and friends for their support, especially, my bright nephew Kaisar Jaxybayev, my fellow student Khaldoon Al-Krad who I have been very lucky to have as my peer in learning, and my friend, Gauhar Serikbayeva, for proofreading my thesis. Finally, I devote this work to my family, Almaz and our daughter Alima. For them I never give up and never stop learning.

I hope this thesis will be a valuable source of information for all its readers. Thank you for your interest in this research!

Abstract

Software developers commonly provide video tutorials for their users on their websites. This study investigated the effectiveness of three practice schedules coupled with instructional videos in software training. For the purposes of the experiments, the instructional videos teaching advanced Word-processing skills were followed by a blocked, interleaved or cumulative practice schedule. The effects of the practice schedules on self-efficacy, engagement, learning, and preference of a practice schedule were compared. Forty-nine university students participated in the experiments. In other domains, contextual interference effect is believed to have caused a difference in influence of practice schedules at least on performance during training and knowledge retention. However, in the current study, no effects of conditions were found on self-efficacy, procedural or conceptual knowledge gain, task performance during training, and on preference of a practice schedule. The theoretical implication of this study is that contextual interference effect in video-based software training should be considered from the perspective of an improved software usability, human-computer interaction, and the video-based nature of instructions. For practitioners, the takeaway is that they can continue using blocked practice as a preferred method of practice in delivering video-based software training.

Key words: contextual interference effect, practice schedule, video-based software training, human-computer interaction.

1. Introduction

Online user training is an essential support service that software developers, such as Microsoft, are expected to provide to users of their product. However, there is a little variation in existing online software user training. Typical software training invites users, first, to watch a video tutorial demonstrating an algorithm for completing a single task or closely related tasks and then to practice the observed task by using several commands dictated by the software interface menu (see Appendix M). To date, there has been little research on the practice approaches in video-based software training. At the same time, research in other domains reveals that sequence (i.e., schedule) of study and practice matters for knowledge durability (see Rohrer & Taylor, 2007; Soderstorm & Bjork, 2015). The current study aims at expanding the scarce research on effectiveness of practice schedules in video-based software training from the users' perspective.

The most common sequence of study and practice utilized in video-based software training reminds of a blocked study and practice approach (see van der Meij & Maseland, 2021). With a blocked practice schedule, a learner practices one type of tasks before moving to the next (Rohrer, Dedrick, and Hartwig, 2020). To compare, with an alternative, interleaved, practice schedule, a learner is encouraged to practice multiple types of tasks in the same session (Rohrer & Taylor, 2007; Dunlovsky, Rawson, Marsh, Nathan, & Willngham, 2013). The latter practice approach has proven to be of a greater benefit for knowledge retention. Yet another alternative, a cumulative practice schedule, requires a learner to practice a newly learned skill along with previously learned skills in a systematic manner (Hughes and Lee, 2019; Mayfield and Chase, 2002). However, effect of a cumulative practice schedule has been under-investigated. Interleaved and cumulative practice schedules are known as mixed practice schedules because, with these practice methods, practice tasks on multiple concepts or categories are mixed within a practice set.

Research reveals that a mixed practice encourages better knowledge transfer and retention over time compared to a blocked practice (Kornell & Bjork, 2008; Kang & Pashler, 2012; Rohrer & Taylor, 2007). The effect is referred to as the contextual interference effect (CI-effect). A blocked practice, on the other hand, results in better performance during practice sessions but is less effective for retention (Soderstorm & Bjork, 2015). Not surprisingly, a mixed practice has been treated as panacea by academia and practitioners. Recent findings, however, prove that the effectiveness of a practice schedule depends on many factors, such as similarity of to-be-learned skills and concepts, the nature of the subject, or the technique used in practice scheduling (see Brunmair & Richter, 2019; Carvalho & Goldstone, 2013; Carpenter & Mueller, 2013).

Unfortunately, there have been only few studies on the effect of practice schedules in video-based software training. For instance, Bouzid and Crawshaw (1987) found that, contrary to massed practice, also known as blocked practice, the so-called distributed practice (i.e., with tasks being spaced over time or separated by unrelated tasks) had a positive influence on the speed and accuracy of performance after Word-processing training. However, the substantially improved usability of software programs could influence the results of investigations (see van Nimwegen, van Oostendorp, Burgos, & Koper, 2006). Therefore, the only recent research on the effect of practice schedules in software training by van der Meij and Maseland (2021) is the most prominent of the kind in the domain as of this date. The investigation compared between the effects of a blocked practice and an interleaved practice in MS Word training on flow, self-efficacy, as well as on performance and learning among primary school students. Notably, van der Meij and Maseland's (2021) expectations of interleaved practice schedule's outperformance were not met. Despite the conclusion that both schedules of practice would be equally effective for software training, van der Meij and Maseland (2021) assumed an impact of the research limitations, such as unavailability of external feedback, on the outcome. Therefore, it is still an open question whether a blocked schedule or a mixed schedule is better for learning software.

The current study will contribute to academia and practice by expanding the research in several ways. First, the study will evaluate rarely investigated effects of practice schedules on learners' engagement, self-efficacy, and preference of a practice schedule along with their influence on task performance and learning in video-based software training. Secondly, the research will measure the understudied effects of a cumulative practice. Third, in the experiments, practice will be preceded by instruction (i.e., video tutorial). Except for the studies of van der Meij and Maseland (2021) and Rohrer and Taylor (2007), in other known studies of this kind, instruction was not part of interventions. Fourth, the van der Meij and Maseland's (2021) assumption that presence of external feedback could result in outperformance of an interleaved practice over blocking will be tested in this study. To accomplish that, this study will allow external feedback in the form of video tutorial replay during practice. Finally, unlike in the van der Meij and Maseland's (2021) study, the investigation will be conducted among university students.

In the next section, first, the theoretical reasoning for the choice of the learning conditions will be presented. Next, research on CI-effect will be described with a detailed review of the causes that contribute or do not contribute to the CI-effect phenomena. Also, human-computer interaction will be discussed as a factor that can influence the effect in software training. Thereafter, the results of the experiments comparing the effects of blocked, interleaved, and cumulative practice arrangements in video-based software training will be reported. Predictions are based on the findings of research on the CI-effect effect; however, domain-specific factors that are likely to influence the study results will be highlighted.

2. Theoretical framework

2.1. Conditions for learning that need to be considered

The effect of presence of instructional support and practice in their various forms on learning has been continuously investigated in efforts to improve learning. Nevertheless, as fairly noted by van der Meij and Maseland (2021), instructional support has been rarely included in the studies of the effect of practice schedules on learning. At the same time, pairing a worked example as an instructional component with practice in the form of problem solving has shown a positive effect on learning among novices (Atkinson, Derry, Renkl, & Wortham, 2000; Sweller, J., van Merrienboer, J., & Paas, F., 1998). Similarly, practice and an instructional component in the form of observation in motor learning showed better transfer results than an experiment with practice only (Shea, Wright, Wulf, and Whitacre, 2000).

2.1.1. Learning benefits of instructional support

One key explanation of the positive effect of instructional support is linked to the cognitive load theory that relates effectiveness of instructional design to the human cognitive architecture (Sweller, van Merrienboer, & Paas, 1998). From the theory perspective, instructional support reduces an extraneous cognitive load. For instance, novice learners, who directly practice problem solving with no instructional support, are forced to select strategies without a prior knowledge and in the absence of an acquired schema (Sweller, van Merrienboer, & Paas, 1998). This process imposes an excessive extraneous cognitive load (Leppink, Paas, van Gog, van der Vleuten, & van Merrienboer, 2014) that can be avoided with inclusion of an instructional support, prior to problem solving.

In software training, users are often offered to observe performance of a task with videos where a model executes the task. However, from the perspective of an excessive cognitive load, instructional video, which is a part of the intervention in the current study, involves a dynamic visualization and, therefore, is believed to impose a cognitive overload due to a continuous information flow (Batrancourt and Tversky, 2000; Tversky et al., 2002). Fortunately, guidelines for designing instructional video (see van der Meij, H., & van der Meij, J., 2013) are aimed at smoothing out the cognitive overload in a video-based instruction. In addition, design of instructional videos has been optimised using a so-called demonstration-based training (DBT) approach. Instructional videos based on the approach aim at facilitating the main processes of observational learning, which are motivation, attention, retention, and production (see van der Meij, H., & van der Meij, J., 2016). The production process is facilitated by including a practice component in training.

2.1.2. What is practice and how much practice is enough?

Next, practice is not merely a “mechanical repetition by rote” (Bernstein as cited by Lee, Swanson, & Hall, 1991, p. 77) but is a problem-solving process. For instance, in motor learning, with each practice trial, a learner does not merely repeat a movement. Instead, the learner rather uses the experience obtained from the previous trial to build a strategy for the movement in the next practice trial (Lee et al., 1991). Therefore, learning how to construct an action plan is vital.

Furthermore, practice is an integral part of demonstration-based training and video-based training is a form of DBT. DBT hinges upon the Bandura’s (1986) theory of observational learning. Practice is aimed at inducing the production process (Grossman, Salas, Pavlas, & Rosen, 2013). Production, in its turn, refers to converting the symbolically represented information into actions, that is, into performance of observed skills (Bandura, 1986). In other words, practice is a mechanism of activation of the observed skills production process.

Nevertheless, there is no sufficient evidence of how much practice is enough. In their review, Soderstorm and Bjork (2015) mention that there is a positive relationship between overlearning (i.e., continuing performance of a task even after achieving 100% accuracy) and long-term retention. Driskell, Willis and Copper (1992) concluded in their meta-analysis that overlearning, that is continuing training even after achievement of the initial proficiency, is beneficial for long-term retention. However, this turned not to be true as the retention interval (i.e., the time lag between the last practice set and the test) increased in a study with learning geography facts and word definitions (Rohrer, Taylor, Pashler, Wixted, & Cepeda, 2005). In a similar way, in the Rohrer and Taylor’s (2007) study with mathematics tasks, practicing nine problems immediately after instruction did not lead to better retention than practicing only three problems immediately after instruction. Thus, more practice is not always an optimal way to learning and might result in a waste of time (Rohrer & Taylor, 2007) and unnecessary costs.

At the same time, Newell and Rosenbloom (1993) concluded that, with an increasing number of practice trials, the task performance time still would keep improving (i.e., decreasing). However, what is important, the rate of improvement would start slowing down at some point of practice trials. One explanation that Newell and Rosenbloom (1993) give to the phenomena is that fast learners could have been effectively completed learning at some point, and, therefore, they do not significantly contribute to performance anymore. In other words, the slowing rate of performance improvement might be a result of slow learners’ performance (Newell & Rosenbloom, 1993). Consequently, more practice does not necessarily contribute to further perfection of a skill or knowledge and, thus, the extra efforts are hardly justifiable.

Furthermore, whether presence of practice per se in training affects learning has also been a subject of disputes in various domains, including video-based software training. For instance, van der Meij and van der Meij (2018) found no effect of the presence of practice on learning in the training where video tutorial was paired with practice. In this connection, the expansion of investigations from the effect of presence of practice per se to the influence of how practice tasks are scheduled on learning could not be overestimated. There exist different practice schedules with varying arrangement of practice tasks in a practice set. The focus of the current study will be blocked, interleaved, and cumulative practice schedules that are discussed in the subsequent sections.

2.2. Blocked, interleaved and cumulative practice

Research has employed a broad variety of practice schedules differing in detail to investigate their effects on learning in various domains. However, it is essential to understand the key features and mechanisms determining success or failure of a particular practice arrangement. In the following sections, the differences in the architectures of blocked, interleaved and cumulative practice schedules will be discussed. In addition, the distinguishing properties and features of the mixed practice schedules, which are believed to cause the contextual interference effect, will be presented.

2.2.1. The structure of a blocked practice schedule

The basic and most common arrangement of practice tasks is a blocked schedule of practice (Rohrer, Dedrick, & Hartwig, 2020). A blocked practice assumes teaching and practicing one concept or skill at a time, that is by blocks. Furthermore, a blocked practice schedule may be massed or spaced. With a massed blocked arrangement, the same type of tasks corresponding to a single topic is practiced one immediately after another, within a single practice set (e.g., Rohrer & Taylor, 2007). The difference of a spaced blocked practice arrangement from a massed blocked practice is that there is an equal time lag between the sets of the same type of practice tasks on the same topic, for instance, 1 week (Rohrer & Taylor, 2007). Another way to space practice tasks is by placing distracting tasks (i.e., not directly related to the concept or a skill being learned) between the focus practice tasks within a single practice trial (Rohrer & Taylor, 2007), in a systematic order.

A massed practice schedule is associated with the overlearning strategy. It relies upon selected findings that practicing the same type of tasks related to the same skill or concept immediately one after another results in better retention. As mentioned earlier, contrary to findings in other domains (see Driskell, Willis, & Copper, 1992), Rohrer and Taylor (2007) found no effect of the degree of massing (i.e., the number of practice problems in a practice set) on learning in mathematics. That is, whether nine or three mathematics problems were practiced in a row after teaching a topic did not influence retention. In addition, a spaced blocked practice outscored the massed practice arrangements by at least 25% in terms of accuracy in the Rohrer and Taylor's (2007) investigation. A practical implication of the finding is that better test results can be achieved solely by a technique, such as spacing of practice tasks, without a need in extra practice (Rohrer & Taylor (2007).

2.2.2. Mixed practice schedules and the contextual interference effect

As opposed to the blocked practice, mixed practice assumes systematic shuffling of practice tasks related to different categories within each practice set. In addition, the tasks could be spaced over several consecutive practice sets. The rationale behind mixing tasks is *the Contextual Interference effect* (the CI-effect), a phenomena initially established in 1966 by Battig in verbal learning. The CI-effect implies that learners test better in knowledge retention and transfer if they practice multiple tasks related to different categories or concepts and mixed within a single practice session. Conversely, retention and transfer performances are worse if tasks are practiced on a single newly learned topic in given practice session. The former is associated with high CI-effect, whereas the latter is related to low CI-effect. Shea and Morgan (1979) further empirically expanded the Battig's discovery to motor learning. Since then, the effect has been repeatedly found in learning artist styles (Kornell and Bjork, 2008; Kang and Pashler, 2012), butterfly and bird species (Birnbaum et al., 2013), mathematics (Rohrer and Taylor, 2007), complex judgement tasks (Helsdingen, van Gog, and Merrienboer, 2011), as well as in other domains.

To the contrary, a mixed practice schedule has shown worse performance during study, at the acquisition phase, compared to a blocked practice schedule (Brady, 1998; Rohrer & Taylor, 2007). One explanation of the reverse effect could be Soderstorm and Bjork's (2015) theory that distinguishes between the retrieval strength and the storage strength of the memory. The former relates to the capacity to retrieve information from the working memory, which is crucial for immediate performance during training. Therefore, it may be assumed that massed practice enhances retrieval strength because a newly learned material of a single type is practiced immediately. On the other hand, based on the Soderstorm & Bjork's (2015) theory, the storage strength accumulates with a mixed practice due to the forgetting effect facilitated by spacing of tasks. The storage strength relates to retrieval of a learned material from the long-term memory, and, therefore, to the delayed performance after training (i.e., learning). Hence, mixed practice does not contribute to immediate retrieval from the working memory because different tasks are practiced within a practice set and forgetting occurs, which could make a mixed practice less beneficial for immediate performance.

2.2.3. The differences between interleaved and cumulative practice schedules

In fact, the term *mixed practice* is an umbrella name for interleaved and cumulative practice schedules. However, there is an important difference between the two practice arrangements. With a cumulative practice (see Table 1), each subsequent practice set systematically includes a task on a skill learned in a preceding lesson, in addition to a task on the newly learned skill. This procedure repeats until all skills have been learned. The practice method assumes a gradual accumulation of all learned skills or pieces of knowledge with every subsequent practice set in a systematic manner (Mayfield & Chase, 2002). Hughes and Lee (2019) note that interleaving by itself assumes a cumulative practice and the only difference lies in how to mix the practice tasks. Unlike a cumulative practice set, however, an interleaved practice set does not necessarily include a task on a previously learned skill. Moreover, a requirement to interleaving is that practice tasks on the same skill component or category should not follow one another (Foster, Mueller, Was, Rawson, & Dunlosky, 2019). The way of mixing practice sets in interleaving are rather decided upon by an instructor. Consequently, only certain arrangements of interleaving could be considered as a cumulative practice.

2.2.4. What makes an interleaved practice more effective for knowledge retention?

One way or another, most studies have compared blocked and interleaved practice. A notable difference of the Rohrer and Taylor's (2007) study from similar investigations is that practice is preceded by instruction. In the study, college students computed the volumes of four geometric solids of different shapes (i.e., a wedge, a spheroid, a spherical cone, and a half-cone) in two experiments. In one condition of Experiment 2, students studied a paper-based tutorial on calculation of the volume of one solid and immediately practiced four problems on computing the volume of that particular solid. Only after finishing with practice on calculating the volume of one solid, they moved to a tutorial and four practice problems on the next solid. This is an example of the earlier defined blocked practice, which prescribes study and practice of one concept or skill at a time. Such an arrangement can also be classified as a massed practice due to no time intervals between practice sets (e.g., in Hughes & Lee, 2019).

Furthermore, in the second condition of the same experiment of the Rohrer and Taylor's (2007) study, students, first, studied the four tutorials each teaching a given solid's volume calculation and then completed all the practice problems in a row. The latter schedule is an example of an interleaved schedule of practice. Rohrer and Taylor (2007) confirmed an impressive finding that an interleaved schedule outperformed a blocked schedule during the delayed posttest. Besides, similarly to the findings in other domains, with a blocked schedule, the accuracy of responses was higher during practice in the Rohrer and Taylor's (2007) study.

Not surprisingly, the interleaving method has immediately grasped attention in the educational world for posttest performance is an indicator of knowledge retention, that is, of learning. Indeed, a low rate of knowledge acquisition during practice could be excused if an interleaved practice definitely ensures retention of knowledge over a longer term (Hughes & Lee, 2019). Nevertheless, based on the literature review, blocking is still a more widespread approach (Rohrer, Dedrick, & Stershic, 2014; Rohrer, Dedrick, & Hartwig, 2020) although it has systematically failed to ensure long-term retention. In this sense, use of the blocked schedule seems to be rather intuitive than based on a conscious, efficiency-driven decision (Dunlosky, Rawson, Marsh, Nathan, & Willingham, 2013). Therefore, numerous attempts to uncover the retention-driving forces behind the magic of interleaving are self-explanatory.

So what is hidden behind such a success of an interleaved practice that makes it superior over blocking when it comes to retention? The studies have agreed that, unlike with a blocked schedule assuming retrieval as a mental activity, an interleaved schedule, in addition, requires discrimination between features of exemplars from different categories. The latter accounts the interleaving effect for the discriminative contrast hypothesis (Birnbaum, Kornell, Bjork, & Bjork, 2013; Kang & Pashler, 2012; Kornell & Bjork, 2008). The attentional bias framework (Carvalho & Goldstone, 2014) further refined the hypothesis**.** In accordance with the framework, interleaving accentuates differences between highly similar categories (Brunmair & Richter, 2019). Further, with interleaving, a learner is induced to select (i.e., to discriminate between) solution strategies for each task. In other words, a learner is imposed to the challenge of matching a taught procedure against a problem to be solved (Bjork and Bjork, 2019; Rohrer & Taylor, 2007). In contrast to a blocked practice, where, by definition, tasks do not highly differ in a practice set, this process requires "an effortful attention" (Hughes & Lee, 2019, p. 418), which brings to the positive effect of interleaving on learning. Thus, the advantage of an interleaved practice is in a greater opportunity to compare solution strategies and identifying distinctive features (Blandin, Proteau, & Alain, 1994).

Along with the *discrimination contrast hypothesis*, literature attributes the interleaving effect to *the distributed-practice hypothesis* (Foster et al., 2019). Whereas *mixing* tasks is a technique associated with the former, s*pacing* is another mechanism contributing to retention (e.g., Rohrer, 2009), in accordance with the distributed-practice hypothesis. Notably, regardless of a schedule of practice, spacing as a technique refers to distributing the same type of tasks related to a given category (Rohrer and Hartwig, 2020). This means that, if, in an interleaved practice set, tasks are sequenced as $P_1P_2P_3$ $P_1P_2P_3$, where P_1, P_2, and P_3 indicate unique problem types, spacing refers to creating a temporal gap between problems of the type P_1 by mixing them with problems of the types P_2 and P_3. If there are more than one interleaved practice sessions in training, this results in additional spacing due to a time lag between the sessions (e.g., $P_1P_2P_3P_1P_2P_3$ – 1 week - $P_1P_2P_3P_1P_2P_3$). By either way of spacing, the positive effect of the approach has been attributed to a reduction of a degree of forgetting because, with each subsequent practice set, a single piece of knowledge is repeatedly refreshed (Blandin, Proteau, & Alain, 1994; Rohrer, 2009).

However, in accordance with research, the contribution of spacing to retention with an interleaved practice has been shown as relative and dependent upon various factors. For instance, as highlighted by Taylor and Rohrer (2010), benefits of spacing could be less apparent in acquisition of conceptual knowledge than with tasks requiring a word-for-word recall only. Secondly, as assumed by Birnbaum et al. (2013), if not applied thoughtfully, spacing could be even harmful for inductive learning, when combined with interleaving, due to interference with discriminative contrast. Third, Taylor and Rohrer (2010) point out that spacing could be more beneficial for retention with a greater retention interval (i.e., the time lag between the last practice task of the final practice trial and the posttest). On the other hand, blocking outperformed interleaving with a relatively short retention interval (Carpenter & Mueller, 2013). Moreover, in Experiment 2 of Birnbaum et al.'s (2013) study on classification of bird and butterfly species, when tasks in a practice set were spaced by mixing with unrelated questions, interleaving did not outscore blocking in retention. And, finally, findings after controlling for the spacing effect still account the benefits of interleaving for the discriminative contrast, rather than for spacing (Taylor & Rohrer, 2010). In other words, the benefit of interleaving on learning seems to be due to the need to discriminate between solution strategies rather than due to a reduced rate of forgetting. Examples of representations of all the aforementioned schedules are presented in Table 1.

Nevertheless, the empirically supported belief in the superior role of the discrimination contrast hypothesis in learning with interleaved practice has been somewhat shaken by Foster et al. (2019). They re-designed the aforementioned Rohrer and Taylor's (2007) experiments with practicing calculation of volumes of three-dimensional solids of different shapes. In Experiment 1 of their study, two more experimental groups were added to the standard blocked and interleaved condition groups initially present in the Rohrer and Taylor's (2007) study. Namely, in the new blocked and interleaved practice schedules only one to-be learned concept (i.e., wedge volume calculation) was retained out of four concepts (i.e., a wedge, a spheroid, a spherical cone, and a half-cone). Besides, the volume calculation tasks were spaced by adding distractive (i.e., unrelated to the focus topic) tasks on permutation and fraction in a single practice session (see Table 2). The aim was to compare effect of interleaving a concept (i.e., wedge volume calculation) with similar concepts (i.e., a wedge, a spheroid, a spherical cone, and a half-cone) and the effect of interleaving the same concept with unrelated concepts of permutation and fraction (i.e., spaced interleaving). In the Foster et al.'s (2019) investigation, the standard interleaving did not outperform the spaced interleaving. This might indicate that spacing does not play a secondary role in the interleaving effect and that a difficult discriminability is unfairly treated as the main factor of greater knowledge retention with mixed practice schedules.

Furthermore, in their systematic literature review, to explain the driving force behind the effect of spacing, Chen, Paas and Sweller (2021) suggest a so-called *cognitive load* hypothesis that hinges upon the principles of the Cognitive Load Theory (Sweller, van Merrienboer, & Paas, 1998). According to Chen et al. (2021), the effect of spacing per se and the effect of spacing with interleaved practice need to be explained by two distinct theories and are of different natures. In particular, when tasks on a concept or a skill are spaced with letting a rest period between the practice tasks or sets, the working memory resource depleted during learning can be recovered during the rest periods (Chen et al., 2021). Consequently, a spaced practice is more effective than a massed practice whereby the recovery of the depleted memory resource is hardly possible by definition. Notably, Chen et al. (2021) do not specifically reject the effect of forgetting on learning that has been commonly associated with the distributed-practice hypothesis. However, unlike the distributed-practice hypothesis, the cognitive load hypothesis proposed by Chen et al. (2021) implies that alternating tasks of different categories in an interleaved practice set does not ensure the spacing effect on learning as such. In accordance with the Chen et al.'s (2021) suggestion, within a single practice set, interleaving of tasks does not assume a recovery of the depleted working memory resource due to the absence of rest periods between tasks. Therefore, from the perspective of the new hypothesis, the driving force behind the positive effect of interleaving on retention is solely the discrimination contrast hypothesis rather than spacing (Chen et al., 2021). To a certain extent, the argumentation breaks the commonly shared assumption that the spacing effect is inherently present with an interleaved practice.

Table 1

Study sequences for various practice schedules

Practice schedule	Practice order
Massed (blocked)	$T_1P_1P_1P_1P_1$- 2 days - $T_2P_2P_2P_2$ P_2 - 2 days - T_3P_3 P_3 P_3P_3 -RI -RT
Spaced (blocked)	T_1 P_1P_1 - 2 days - P_1P_1– 2 days - $T_2P_2P_2$ - 2 days - P_2P_2– RI – RT
Interleaved	$T_1T_2T_3P_1P_2P_3P_2P_1P_3$ – RI – RT
Interleaved (two sessions)	$T_1T_2T_3P_1P_2P_3$ - 2 days - $P_2P_1P_3$ – RI – RT
Interleaved with distractive tasks	$T_1T_2T_3P_1D_1D_2$ $P_2D_1D_2P_3$ – RI – RT
Cumulative	T_1P_1– 2 days – $T_2P_1P_2$– 2 days – $T_3P_1P_2P_3$– RI – RT

Note. T Tutorial (e.g., T_1– tutorial teaching skill component 1), *P* Practice task (e.g., P_1 – practice task on skill component 1), *RI* Retention Interval (i.e., a time gap between the last practice task and retention test), *RT* Retention Test, *D* Distractive Task (e.g., D_1 – a task or question not directly related to the skill being learned).

Table 2

Practice orders in the studies of Rohrer & Taylor (2007) and Foster et al. (2019)

Practice schedule	Practice order
Rohrer & Taylor (2007):	
Massed (blocked)	*session 1*: $T_1P_1P_1P_1P_1T_2P_2P_2P_2P_2T_3P_3P_3P_3P_3T_4P_4P_4P_4P_4$
	– 1 week rest –
	session 2: $T_1P_1P_1P_1P_1T_2P_2P_2P_2P_2T_3P_3P_3P_3P_3T_4P_4P_4P_4P_4$
Interleaved	*session 1*: $T_1T_2T_3T_4P_1P_2P_3P_4P_2P_3P_4P_1P_3P_4P_1P_2P_4P_1P_2P_3$
	– 1 week rest –
	session 2: $T_1T_2T_3T_4P_2P_3P_4P_1P_3P_4P_1P_2$ $P_1P_2P_3P_4P_4P_1P_2P_3$
Foster et al. (2019), Experiment 1:	
Massed (blocked)	$T_1P_1P_1P_1P_1$ $T_2P_2P_2P_2$ P_2 $T_3P_3P_3P_3P_3$ $T_4P_4P_4P_4P_4$
Interleaved	$T_1T_2T_3T_4P_1P_2P_3P_4$ $P_1P_2P_3P_4P_1P_2P_3P_4P_1P_2P_3P_4$
Interleaved (spaced with distractive tasks)	$T_1T_{D1}T_{D2}T_{D3}P_1P_{D1}P_{D2}P_{D3}P_1P_{D1}P_{D2}P_{D3}$ $P_1P_{D1}P_{D2}P_{D3}P_1P_{D1}P_{D2}P_{D3}$

Note. T Tutorial (e.g., T_1 – tutorial teaching skill component 1, T_{D1} – tutorial on distractive task 1), *P* Practice task (e.g., P_1 – practice task on skill component 1, P_{D1} – practice task on distractive task 1).

2.2.5. A cumulative practice versus an interleaved practice

While the scientific world has not yet agreed on what makes mixed practice so beneficial for learning, in general, the difference between blocking and mixed practice is clear enough. However, the question on what makes a cumulative practice conceptually different from interleaving remains to be answered. Taking into account mixing of tasks in both schedules, definitely, cumulative practice capitalises on the benefits of the discrimination contrast hypothesis similarly to interleaving. Moreover, the classic assumption of the benefits of spacing cannot be accounted for uniqueness of a cumulative practice either, because the assumption is equally valid with respect to interleaving. For the same reason, the recently suggested Chen et al's (2021) cognitive load hypothesis can hardly serve as an argument to show a cumulative practice distinct from interleaving. Indeed, with both practice schedules, the depleted working memory resource is supposed to be restored in the rest periods between practice sessions, thus, contributing to knowledge retention. Hence, it is difficult to judge what can make cumulative practice more or less effective compared to interleaving.

Nevertheless, a cumulative sequence is unique in that it prevents "teaching and leaving" before mastering a skill to "initial mastery criteration" (Hughes and Lee, 2019, p. 416). That is, a learner is supposed to master one skill before stepping into another practice session where a new skill practice is added to practicing the previously mastered skill. In other words, with a cumulative schedule of practice, a certain degree of blocking is permanently present within each intermixed practice set until a given skill has been mastered. As Hughes and Lee (2019) fairly note, such discrete blocking can negatively affect the benefits of mixed practice for learning. For comparison, with an interleaved schedule, tasks on different component skills can be mixed randomly at the discretion of an instructor and blocking can be avoided. Consequently, cumulative practice needs to be chosen with a caution wherever blocking can hinder knowledge retention.

At the same time, Mayfield and Chase (2002) attempted to decompose the architecture of cumulative practice into the three elements inherent in the practice technique. Namely, the elements are (1) mixing practice tasks that results in a higher retention rate due to the discrimination contrast benefits, (2) distributed practice that results in the spacing effect, and, finally, (3) an extra practice. The aim of the study was to investigate which of the elements would be crucial for application (i.e., transfer), problem solving, and posttest on retention by comparing the effects of the three conditions on learning algebra rules. The experiment conditions were a cumulative practice, a review (i.e., a practice spaced over time without cumulating practice tasks), and an extra practice (i.e., an additional round of practice in a blocked mode). Although no difference was found in the effect of practice schedules on retention, cumulative practice was superior over the two other conditions in terms of an average proportion of correct responses across interim tests executed between practice sets. The Mayfield and Chase's (2002) interpretation of the results was that discrimination contrast hypothesis, and not extra practice or spacing, was the key driving force for the positive effect of a cumulative practice in the interim tests. At the same time, Mayfield and Chase (2002) related the absence of the cumulative practice's effect on retention to an increasing turnover of participants in groups and a lack of the statistical power due to a large time lag between training and the posttest. In other words, the result was not anyhow related to a possible inadequacy of a cumulative practice for retention. Thus, there are questions to be answered yet. The contradictions in findings pointed out in the next section further escalate a degree of curiosity around the effectiveness of various practice schedules.

2.3. Is mixed practice always good for learning?

There have been multiple propositions to foster the idea of the mixed (e.g., interleaved and cumulative) practice schedules in learning since discovering the phenomena of interleaving. Indeed, the scientifically backed temptation to apply the mixed schedules of practice, typically, an interleaved arrangement, wherever it is possible, is justifiable based on multiple scientific evidence. Nevertheless, in particular cases, a success of the mixed schedules of practice seems to be subject to several specificities. They include but are not limited to learners' prior knowledge, similarity of to-be-learned categories, the freedom to choose a schedule by learners, the learning material, a type of posttest, and to which extend an inductive learning can be facilitated with respect to a certain context. Thus, generalization of the findings could be premature.

For instance, Rau, Aleven, and Rummel (2010) studied the effect of a blocked versus an interleaved practice with multiple graphical representations of fractions among 5 and 6th grade students. Contrary to the robust findings in mathematics, the research showed no advantage of an interleaved practice on the representational knowledge (i.e., "the ability to interpret representations of fractions and to use them to make sense of fractions"; Rau et al., 2010, p.417) and the operational knowledge (i.e., the ability to solve problems on fractions) during immediate and delayed posttests. It needs to be mentioned though, that Rau et al. (2010) related the contradictory finding to the design of the study and to the participants' representational fluency existing prior to the experiment.

2.3.1. The category discriminability and the contextual interference effect

As aforementioned, one of the well-established key explanations of the interleaving phenomena is discrimination contrast hypothesis. In other words, drawing comparisons between different strategies during practice is believed to lead to long-term retention. However, Higgins and

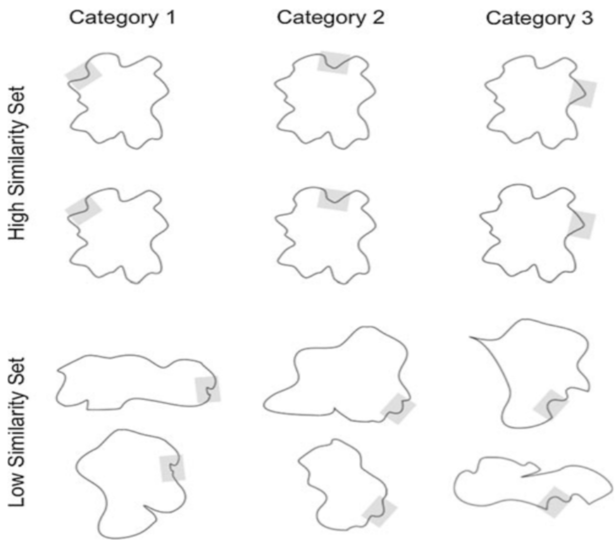
Ross' (2011) study demonstrated that comparison of categories did not necessarily bring to a better ability to classify a stimuli into a category. The research reveals that performance depends on the type of comparison (e.g., between-category or within category) induced by practice.

At the same time, the difference between within-category and between-category comparison might seem confusing at the first glance. The difference is illustrated in Figure 1 that reproduces an image of exemplars used in the study of Carvalho and Goldstone (2014). In their study, one set included images belonging to different categories that, however, were highly similar to each other within and between the categories (i.e., high-similarity set). That is, stimuli from category 1 in the set were highly similar to other stimuli from the same category as well as to stimuli from a different category of the same set. To the opposite, the second set included exemplars being highly dissimilar to each other within and between the categories (i.e., low-similarity set). In Experiment 1 of the Carvalho and Goldstone's (2014) study learners studied the categories from either of the sets both, in a blocked and in an interleaved condition. The aim of the experiment was to investigate whether the effectiveness of a practice schedule would be influenced by a to-be-learned category structure during and after learning activities (i.e., performance of study tasks and generalization tasks). In Experiment 1, Carvalho and Goldstone's (2014) found that when both, within and between-category similarities were low (i.e., low similarity categories), a blocked practice resulted in a better ability to classify a presented stimuli into one of the categories during training (i.e., during performance of study tasks) than interleaving. However, performance after training did not differ between the blocked practice and the interleaved practice conditions. Carvalho and Goldstone (2014) concluded that, in category learning, success of an interleaved practice depends on the category structure.

Overall, a key conclusion of Carvalho and Goldstone (2014) was that, whereas the challenge to find differences between categories is responsible for the effect of an interleaved practice, the challenge to find commonalities within low-similarity categories results in an improved performance with blocking. Hence, in light of the finding, it is possible that manipulations with a category structure applied with an appropriate practice schedule for a given context could compound the effect of a practice schedule on learning. Another interesting suggestion of Carvalho and Goldstone (2014) is that blocking could be more beneficial for learning a novel material. On the other hand, earlier, based on their findings in the domain of algebra, Mayfield and Chase (2002) suggested using a cumulative practice for learning novel material. Whether the suggestions are a mere coincidence or could be related to the architectures of the two practice schedules is an intriguing question.

Figure 1

High and Low Similarity Category Sets



Note. The upper part represents a study set of exemplars from categories with a high similarity. The lower part of the picture represents a study set of exemplars from categories with a low similarity. The shaded areas indicate the common distinctive feature for each category. (2014). "Putting category learning in order: category structure and temporal arrangement affect the benefit of interleaved over blocked study", by P. Carvalho, and R. Goldstone, 2014, *Memory & Cognition, 42*(3) (http://doi.org/10.3758/s13421-013-0371-0). Copyright 2013 by Psychonomic Society, Inc.

Furthermore, Carpenter & Mueller (2013) have explicitly demonstrated that success of a practice schedule depends on the degree of the between-category discriminability in their investigation with learning French pronunciation by English native speakers. In the multiple experiments of the study, participants of the blocked practice condition systematically showed better results in associating pronunciation rules with words (e.g., the rule of pronouncing the letter combination "ou" with words mouton, genou, verrou) than participants of the interleaved condition (Carpenter & Mueller, 2013). What is important, the result was consistent regardless of the level of exposure to the material, that is, regardless of the number of words learned per a rule (e.g., four or fifteen words; Carpenter & Mueller, 2013). In addition, in the Carpenter and Mueller's (2013) multiple-experiment study, whether participants were informed of existence of the pronunciation rules for to-be-learned words did not reverse the effect of blocking. Changing a form of testing from multiple choice to recall tests did not reverse it, too. Such a consistency points at the relative robustness of the findings.

In accordance with Carpenter and Mueller (2013), one explanation to such a disagreement with the earlier established success of an interleaved schedule would be that the interleaved practice sets included words with highly discriminable pronunciation rules (e.g., eau, ch, s, t). The distinguishability enabled participants to easily recognise which rule to apply for a particular word (Suzuki, Yokosawa, & Aline, 2020). Thus, a desirable difficulty (Bjork & Bjork, 2020), that ensures the postponed interleaving effect, was not facilitated. Indeed, the more recent research of Brunmair and Richter (2019) reveals that the interleaving effect is more apparent when categories are similar to each other. The Carpenter and Mueller's (2013) finding draws attention back to the Carvalho and Goldstone's (2014) explanation that, with highly dissimilar categories, a better performance with a blocked practice could be driven by the challenge to find commonalities, rather than differences, between the categories. Carpenter and Mueller (2013) also assumed that one of the reasons for the positive effect of blocking in their study could be the need to identify commonalities in the pronunciation rules. Consequently, differences and commonalities in a category structure need to be taken into account in training a target skill.

2.3.2. Boundary conditions for presence of the contextual interference effect

Studies, such as of Carvalho, Braithwaite, de Leeuw, Motz, and Goldstone (2016), show that superiority of the interleaved method over a blocked schedule is not a decided fact. In their experiment conducted in an in-vivo condition (e.g., not a laboratory condition), they established that, when deliberately chosen by learners as a method of study, a blocked schedule outperformed interleaving. Therefore, success of a study sequence also seems to be dependent on whether a practice schedule is deliberately chosen by a learner or imposed on him or her. Similarly, in the opinion of the participants of the Carpenter and Mueller's (2013) study, blocking lead them to a better pronunciation of French words. However, earlier, Birnbaum et al. (2013) warranted that preference of blocking by a learner would not necessarily lead to learning. Learners tend to confuse performance during training with learning, which may result in choosing an inappropriate method of learning (Soderstorm & Bjork, 2015). Therefore, a sequence of practice needs to be chosen with a caution, because blocking may be a strategy preferred by learners due to the immediate positive results during practice whereas, ultimately, it can be not effective for learning. This illusionary success of blocking, wherever it has been proven not to be actually effective, could be a consequence of the disappointing results of interleaving during the study phase. On the other hand, due to the relative consistency of the findings on the benefits of interleaving, blocking could be unfairly ignored despite a potentially positive learning outcome.

Furthermore, outperformance of blocking during practice only but not in posttests has served as a central pro argumentation favouring the mixed schedules of practice. Literature reveals that a successful performance during practice does not necessarily lead to learning (Soderstorm & Bjork, 2015; Rohrer & Taylor, 2007). At the same time, posttest scores, and not scores during practice, indicate the presence of long-term retention and transfer, which constitute learning and imply durability of knowledge (Soderstorm & Bjork, 2015). Nevertheless, with their multiple-experiment study, Carvalho and Goldstone (2020) contributed to the controversy around the advantages of an interleaved sequence. They showed that, with certain testing methods, such as writing definitions, the blocked condition outperformed the interleaved condition in posttest. A key explanation to the sub-standard finding is that studying with a blocked mode contributes to an autonomous representation of learned material in a learner's cognitive processing. Therefore, as writing a definition is aimed at testing the "isolated and independent knowledge" (Carvalho & Goldstone, 2020, p. 84), a blocked schedule is more effective for retention of definitions, if the definitions are dissimilar (Carvalho and Goldstone, 2020). To the contrary, with multiple-choice questions, the discriminative contrast comes into play if learned concepts are hardly discriminable (see Birnbaum, Kornell, Bjork, & Bjork, 2013; Carvalho & Goldstone, 2014; Kornell & Bjork, 2008; Kang & Pashler, 2012). Therefore, if the latter discriminability condition is fulfilled, the testing method will demonstrate better retention with interleaving than with blocking (Carvalho and Goldstone, 2020). A possible influence of a testing method was earlier noted by Birnbaum et al. (2013) who assumed that interleaving would show positive effects if a posttest called for discrimination between categories.

In their multi-level meta-analysis of the interleaving effect, Brunmair and Richter (2019) studied its generalizability to different settings and learning materials. The results varied from a high effect of interleaving in studies with painting to no effect in studies with expository texts and tastes. This implies that the effectiveness of an interleaved practice can differ based on a setting and a type of a learning material, too. Hence, it should be generalised with a caution (Brunmair & Richter, 2019). In addition, it seems that the interleaving effect is a derivative of concrete design factors, such as implementation, characteristics of learning materials, similarity of categories (Nemeth et al., 2019).

Importantly, an interleaved condition did not outperform a blocked condition in software training, too, in either of the testing moments (i.e., during practice, immediate and delayed test; see van der Meij & Maseland, 2021). However, it is fair to mention that van der Meij & Maseland (2021) are far from considering the outcome of the experiment to be conclusive. The deviation from the common trend has rather been explained by the limited number of practice trials allowed in the experiment. In addition, absence of an external feedback (e.g., checking correctness of a response by re-playing a video tutorial) during training could have diminished the CI-effect (see van der Meij & Maseland, 2021). Van der Meij and Maseland (2021) argue that, with interleaving, a temporal gap between tasks related to the same topic or concept creates a greater chance of forgetting a previously learned material than with blocking. In van der Meij and Maseland's (2021) opinion, external feedback could be a boundary condition for the presence of the CI-effect. In fact, Rohrer (2019) emphasizes that forgetting of an acquired knowledge, and not necessarily poor acquisition of knowledge, could be a reason for a low retention rate. Therefore, in accordance with van der Meij and Maseland (2021), presumably, feedback, such as replaying a video instruction during practice, would enhance retention with interleaving. Eventually, van der Meij and Maseland (2021) concluded that both practice sequences would be suitable for software training.

Finally, interleaving has shown to be especially effective in inductive learning. Induction assumes that learners derive rules and patterns through studying exemplars of a category or a concept (Carvalho & Goldstone, 2014). Inductive learning assumes active learning through inquiry. One well-proven example of successful use of interleaving in inductive learning was categorization of works of different artists by painting styles through viewing their works in various sequences as it was experimented by Kornell and Bjork (2008). The study was later replicated by Kang and Pashler (2012). On the other hand, non-inductive (i.e., deductive or rule-based) learning is different from inductive learning in that ready-to-use rules are initially shared with learners without letting them discover concepts on their own. For instance, if learners were initially given a certain rule to match a painting style against an artist's piece of work, this would be an example of non-inductive learning. Noh, Yan, Bjork, and Maddox (2016) established that blocking lead to a better accuracy than interleaving in a categorization posttest if, during the study phase, categorisation rules were presented in a more explicit manner. In other words, learners were able to more accurately attribute items to an appropriate category if, during the study phase, they were clearly informed of what features make an item referring to a certain category. At the same time, the study demonstrated that, when learners were informed of the categorization rules less explicitly, interleaving was superior over blocking in a posttest.

On the other hand, in the aforementioned Carpenter and Mueller's (2013) study, blocking was more favourable than interleaving with and without making participants aware about existence of the pronunciation rules for to-be-learned French words prior to practice. One explanation would be that the dissimilarities in painting works in Kornell and Bjork's (2008) and Kang and Pashler's (2012) studies could be more subtle between and within the categories, whereas the differences in the pronunciation rules would be easier to notice. Nevertheless, based on their findings, Noh et al.

(2016) suggest that, wherever the categorization rules can be verbalized less explicitly (e.g., in learning painting styles), blocking could be more beneficial for learning. And vice-versa, in certain science domains, such as chemistry or mathematics, where item classification rules could be made more explicit, interleaving would be more practical (Noh et al., 2016). However, in the real life, the borderline between a rule-based and a non-rule-based learning can be subtle (Noh et al.,2016).

Overall, a large body of literature has agreed upon one-sided conclusions on always-winning interleaved sequence of practice. Nevertheless, there is a growing number of scientific evidence that does not fit in the common picture. They reveal that either of the practice schedules can be advantageous in terms of learning outcomes. Selective research even suggests blocking at the initial stage of knowledge acquisition to reach automation and switch to a spaced and interleaved mixed practice at a later stage (Carpenter & Mueller, 2013; Firth, 2018). Thus, instead of stereotyping advantages of a mixed practice, choice of a practice sequence needs to be guided by specificities of a domain and various conditions (e.g., contextual or testing) that might wipe out a positive effect of either of a practice and a study schedule (see Carvalho & Goldstone, 2019; Suzuki, Yokosawa, & Aline, 2020). Secondly, prior to abandoning the straightforward blocked method of study, it needs to be assessed whether in a particular context it is feasible to create the desirable difficulties (Bjork & Bjork, 2020) claimed to be a key success factor with a mixed practice.

2.4. Practice schedules and the human-computer interaction

In light of the mixed findings, the issue of generalizability of the positive findings to software training that involves the human-computer interaction (HCI) appears to be even more intriguing. Indeed, peculiarity of the domain dictates a thorough consideration of its distinctive features. Definitely, the recent study of van der Meij and Maseland (2021) is a breakthrough because it has expanded the investigation of the kind to the software-training domain. However, unfortunately, there seem to be no published research linking particularly peculiarities of HCI to the CI-effect and, consequently, to schedules of practice. Partially, such a scarcity of research in the domain can be explained by challenges of externalizing the cognitive processes underlying the HCI. The only recent aforementioned study in the domain conducted by van der Meij and Maseland (2021) related the inconsistent finding to the low number of practice trials and non-presence of a feedback during practice. However, a deeper insight into the nature of the HCI could raise additional accounts for placing a blocked practice at the same level with a more sophisticated mixed practice, with respect to software training.

In fact, it is hardy disputable that, compared to other domains, the human-computer interaction is an interdisciplinary domain found at the intersection of cognitive psychology, behavioural science, computer science, sociology, and even anthropology (Waddell, Zhang, & Sundar, 2015). Sundar and Nass (2010) concluded that computer was an independent source of information by itself. In the human-computer interaction, just like in instructional design, the Cognitive Load Theory served as a foundation to reproduce the human cognition and increase the memory capacity of a user (Hollender et al., 2010). As it is fairly noted, computers are not only information processors, but also are the "active agents capable of learning and solving problems" (Waddell et al., 2015, p. 2). Zhang and Norman (1994) have accurately defined HCI as " the interwoven processing of internal and external information that generates much of a person's intelligence" (Zhang & Norman, 1994, p. 87). This is what distinguishes HCI from human cognition, whereas the latter explains information processing "exclusively inside the mind of one person" (Rogers, 2004, p. 106). Therefore, a possible CI-effect in software training has to be hypothesised based on the properties of the human-computer interaction.

First, as a matter of fact, with software, an instructional support continues even after delivery of an instruction (e.g., after instructional video or study of a manual) due to presence of the interactive menu. Indeed, menu is, in a way, a procedural scaffolding tool (i.e., a guidance on how to use resources (Huang, H-W., Wu, Ch-W., & Chen, N-Sh., 2012), which is accessible by a user at any time during practice. Besides, in software training, software itself provides immediate feedback on whether a particular manipulation has been correctly completed (van der Meij & Maseland, 2021). In other words, paradoxically, in the human-computer interaction, a computer plays the role of an expert assisting a human. Inevitably, while being a product of the human cognition, a computer influences a user's cognitive load.

Secondly, usability is an ultimate goal in the human-computer interaction (Hollender, N., Hofmann, C., Deneke, M., & Schmitz, B., 2010). At the same time, Bjork & Bjork (2011) define "creating desirable difficulties to enhance learning" (p. 58) as a key success factor of interleaving and spacing of practice tasks. However, with the tremendous usability improvements since emergence of computers, it is hardly achievable to create conditions for strategy discrimination in software practice. The simplest example is that spelling error prevention is not anymore a function of typewriting appropriate commands but just an issue of selecting an option in a menu (Holleder et al., 2010). In other words, in contrast to the learning technology, productivity tools, such as Microsoft Word, are designed to have a task completed in the most efficient way (Tseopis, Avouris, & Komis, 2008). However, the improved usability might not be beneficial for learning (Holleder et al., 2010). Therefore, usability and effective learning seem to be conflicting when a productivity tool is used as a learning tool itself.

van Nimwegen et al. (2006) argue that, with a greater reliance on the interface, learners are not induced to think more and to store information in their memory. The usability goal is to reduce a memory load by having users recognize rather than recall with the help of externalization of information on the interface (Holleder et al, 2010). van Nimwegen et al. (2006) suggested that “less assisted” human-computer interactions would induce better planning, engagement and performance of users. Furthermore, the same study implies that, from the perspective of the cognitive load theory, making use of software more effortful is comparable with increasing germane cognitive load. The latter involves construction of schemas and, therefore, is important for learning to take place. Thus, usability is achieved at the cost of the learning performance (van Nimwegen et al., 2006).

Furthermore, recent research proposed to sophisticate discrimination of solution strategies even further with the goal to reinforce the effect of interleaving. Interleaving is more effective when practice sets include a mixture of problems from similar categories, as opposed to unrelated categories (Firth, 2018). The explanation is that discrimination between items from similar categories is subtle (Kornell and Bjork, 2008). Earlier Rohrer, Dedrick, and Burgess (2014) also found a positive effect with a more intense problem discrimination and association condition in mathematics problem. However, with software, solution strategies are already built-in in the programs with a reverse aim to limit possibilities for errors. Hence, there seem to be a little chance for a learner to face a sophisticated strategy selection scenario. Nevertheless, the exception could be complex programs, whereby outcome may still primarily depend on human cognition (e.g., engineering).

In addition, it is well established, that limitations of the human working memory need to be facilitated for an effective learning (Sweller, van Merienboer, & Paas, 1998). Miller (1956) proposed to avoid overloading the working memory by organising information into a sequence of chunks. He claimed that the working memory is capable of holding, on average, seven elements of information. Indeed, the way commands for each operation are conveniently grouped in the word processor menu must be an echo of the Miller’s “magical number 7” theory (Rogers, 2004). Thus, with the improved interface design, the working memory is released from processing of a part of information. In other words, the rationality of using more sophisticated practice schedules should be weighed against an inevitable cognitive load reduction due to the rapid technological development.

Finally, in their attempt to integrate the cognitive load theory and concepts of the human-computer interaction, Holleder et al. (2010) classify vocabulary learning as learning with a low intrinsic cognitive load (i.e., a material-complexity dependent load), as opposed to learning to construct sentences. In the former, each word can be learned independently of each other. Similarly, in math, a prior knowledge of the concept of fractions is required to acquire knowledge on fractions multiplication. However, in word processing, operations are fairly independent of each other, that is, adjusting margins of the text can be learned without knowing how to use the AutoText function. Overall, learning word processing does not seem to require a complex schema construction associated with strategy discrimination.

2.5. Dependent variables

2.5.1. Self-efficacy

Intriguingly enough, presence of an acquired skill does not ensure achievement of an expected outcome unless a person perceives him- or herself of a high self-efficacy (Bandura, 1986). Bandura (1986) defined perceived self-efficacy as one's judgement of own "capabilities to organise and execute courses of action" (page 391) to achieve a planned level of performance. He distinguished between self-efficacy and the expected outcome that is a consequence of an action. In addition, in accordance with Bandura (1986), people judge of their self-efficacy not only based on their own enactive experience. Observation of others executing a task successfully can contribute to self-efficacy (Bandura, 1986, p. 399).

Nevertheless, with respect to software training, van der Meij and Maseland (2021) found no significant difference in the effect of blocked and interleaved practice schedules in software training on self-efficacy improvement. However, the finding was not related to structural features of the schedules. Such a finding was rather accounted for the confidence of participants in both experimental conditions gained through task completion during training (van der Meij & Maseland, 2021). Therefore, conclusions on the effect of particular practice sequences on self-efficacy seem to be pre-mature.

2.5.2. Student engagement

Trowler's (2010) literature review is arguably a multifaceted analysis of existing interpretations of the concept of student engagement. There is seems to be no unique definition of engagement to this date due to impossibility to measure engagement directly. However, one aspect of engagement that has proven to highly correlate with educational outcomes is a learners' time investment. The greater time learners invest to study, practice, problem solving, the more they learn (Kuh, 2003). Guo, Kim, and Robin (2014) used time people watched instructional videos as an aspect to judge learners' engagement. Compared to that, the current study focuses on the effect of a schedule of practice paired with instructional video on educational outcomes, rather than on the effect of instructional video only. However, time invested by a learner in watching video can serve as an indication of students' engagement in the video-based training as a whole (i.e., instructional video and practice). Besides, it can be an important indicator of the feedback frequency, which could affect retention (see Schmidt and Bjork, 1992).

2.5.3. Learning

The effect on learning in different conditions is assessed with measuring procedural knowledge and conceptual knowledge, also referred to as declarative knowledge. The borderline between conceptual and procedural knowledge seems to be subtle in software training. Declarative or, conceptual, knowledge is about “facts and things” (Anderson, 1985, as cited by Yi and Davis, 2003), whereas procedural knowledge is about how to perform an activity. In learning cognitive skills, procedural knowledge builds upon declarative knowledge through knowledge compilation (Yi & Davis, 2003). Through compilation, learners acquire procedural skills by integrating “the sequences of cognitive and motor processes required to perform the task” (Kanfer & Ackerman 1989, p. 660, as cited by Yi & Davis, 2003). Yi and Davis (2003) conclude that declarative knowledge is a pre-requisite to the skilled task performance, subsequently, to exercising procedural knowledge. Learners observe task performance, then the task rules are encoded and stored, and, finally, learners are able to pick a strategy to complete a task.

However, cognitive processes in learning word processing skills are different from those, for instance, in math problem solving. Whereas in the latter the process of schema construction is more sophisticated, in software training, due to the presence of the improved menu and the interface, lower efforts are required for constructing declarative knowledge. With Word-processing, the required conceptual knowledge rather refers to being able to visually associate certain tasks with corresponding menu options. Therefore, conceptual and procedural knowledge in software training need to be measured in conjunction with each other.

2.5.4. The learners’ preference of a schedule of practice

Research reveals that learners overwhelmingly judge blocking as a more effective schedule of study in the experiments whereby interleaving has obviously been more effective for learning (Yan, Bjork, and Bjork, 2016; Yan and Soderstrom, 2017; Zulkiply and Burt, 2013). Interestingly, even awareness of a greater effectiveness of interleaving, compared to blocking, has not reversed learners’ judgement (see Yan et al., 2016). The “metacognitive illusion” (Yan et al., 2016, page 918) has been explained with the perceived sense of ease while studying in a blocked manner (Kornell and Bjork, 2008). The sense of ease, in its turn, might be stimulated by the feeling of fluency that learners think they gain because exemplars related to the same category or the same concept have been presented one right after another (Kornell & Bjork, 2008; Yan et al., 2016). At the same time, the learners’ judgement seems to affect their preference of one way of studying over the others. At the same time, the resulting choice of the study method taken in accordance with a learner’s personal preference may not be the most efficient as studies have shown.

Nevertheless, the aforementioned studies of learners’ preferences have been conducted in the domain of category learning whereby interleaving has consistently shown a superior effect over blocking, unlike in the video-based software training (see van der Meij & Maseland, 2021). In other words, it is questionable whether the preference findings established in other domains would be replicated with the video-based software training. Indeed, the increased software usability might supposedly neutralise “the desirable difficulties” (see Bjork & Bjork, 2011) caused by interleaving and make learners’ preference less sensitive to any particular study schedule.

3. Research design and research questions

The current study aims at investigating the influence of practice schedules on user affect, behaviour and cognition in video-based software training. The training will include instructional videos demonstrating performance of six Word-processing tasks. This study has an experimental design and includes three conditions: a blocked practice, an interleaved practice, and a cumulative practice (please see Appendix B for the practice arrangements). The following research questions were investigated:

Research question 1 (RQ1): what is the effect of a practice schedule on self-efficacy?

The most recent research shows that, if tasks are well practiced during software training, self-efficacy increases, regardless of a practice schedule employed (van der Meij & Maseland, 2021). In the current study, too, tasks will be practiced repeatedly with all the three practice schedules. However, with a cumulative practice schedule, combination of the blocking elements with "the desirable difficulties" is likely to increase the level of self-efficacy due to the perceived sense of confidence gained through both, the consistency in presenting the tasks and mixing. Therefore, it is expected that, while all the three practice conditions will positively affect learners' self-efficacy, the effect after training will be greater in the cumulative practice condition than in the blocked or interleaved conditions.

Research question 2 (RQ2): what is the effect of a practice schedule on learners' engagement (i.e., viewing of the instructional videos)?

The following measures can serve as a proxy for engagement in video-based software training: total playtime, unique playtime, and replay time (see the Method section for definitions). Although total playtime shows total time invested in viewing a video or videos and does not carry any information about quality of viewing. Contrary to that, unique playtime shows whether a taught material has been covered in full. Replay time, in its turn, indicates a feedback frequency, which is believed to affect retention (see Schmidt and Bjork, 1992). Consequently, the two latter measures are more informative measures of engagement as a dependent variable.

With a blocked and a cumulative practice, learners are imposed to a lower degree of forgetting than with interleaving as a given task is repeatedly practiced in a consistent manner in each practice set. As a result, speaking of replay as a measure of engagement, in contrast to a interleaved practice, learners in the two other practice conditions might use the feedback possibility less intensively. In a similar way, it is likely that a perceived sense of ease that might be gained with a blocked and a cumulative practice at some point would result in skipping larger parts of instructional videos in the two conditions. By definition, skipping parts of videos would lead to a lower unique playtime. Therefore, it is expected that, for an interleaved schedule of practice, learners' engagement in terms of both, video replay and unique playtime, will be greater compared to a blocked or a cumulative practice.

Research question 3 (RQ3): what are the effects of a schedule of practice on task performance during video-based software training and on learning (i.e., conceptual and procedural knowledge gain after training)?

The study compares the effects of the three practice schedules on the three dependent variables: task performance during training (i.e., at the knowledge acquisition stage), conceptual knowledge gain after training, and procedural knowledge gain after training. It is worth noting that the measurement of conceptual knowledge will give a more complete understanding of the effects on learning compared to the latest research of this kind where the effect on conceptual knowledge was not evaluated (e.g., van der Meij & Maseland, 2021). The aiding properties of the interface menu, the video replay possibility at the acquisition stage, and a high discriminability of the task performance algorithms in MS Word are likely to challenge the established power of the CI-effect. Therefore, first, it is expected that, during training (i.e., at the knowledge acquisition stage), a blocked practice will not yield greater task performance scores than an interleaved or a cumulative practice schedules. A second prediction is that interleaved and cumulative practice schedules will not lead to a higher conceptual or procedural knowledge gain than a blocked schedule of practice. Third, it is expected that an interleaved practice will not lead to a higher conceptual or procedural knowledge gain than a cumulative practice.

Research question 4 (RQ4): what is the effect of a schedule of practice on the learners' preference of a schedule of practice in video-based software training?

Assumption that blocking of practice tasks could be associated with the feeling of ease and fluency among learners is likely to prove valid with video-based software training, too. Presumably, the factors will direct preference of learners towards learning with a blocked practice method, regardless of what the learning outcome is. Therefore, it is expected that, with MS Word training, a blocked practice will be preferred over mixed practice schedules. At the same time, for the same reason, it is expected that a cumulative practice will be preferred over an interleaved practice schedule.

4. Method

4.1. Participants

Fifty-two university students participated in this study. Forty-nine students were included in the study. Three students were excluded due to not complying with the participation requirements resulting in sixteen participants in the interleaved condition, sixteen participants in the blocked condition, and seventeen participants in the cumulative practice condition. 83.67 % of the participants were between 18 and 25 years old, 8.16 % were between 26 and 35 year old, and 4.08% of participants were over 35 years old. 71.43 % of participants reported themselves as bachelor students, 16.33% as master students and 12.24% as having "other" student status. There were twenty-eight male participants (57.14%) and twenty-one female participants (42.86%). All participants were users of Windows-system and Microsoft Word versions of 2016 or later.

The students were randomly assigned to the conditions. The blocked practice group included nine male students and seven female students; the interleaved practice group included twelve male students and four female students; the cumulative practice group included seven male students and ten female students. Approval for the study was obtained from the Ethical Committee of the University of Twente.

4.2. Materials

4.2.1. Instructional videos

The video training material was segmented into one introductory and 6 short video tutorials teaching one Word-processing skill each. The segmentation approach ensured presentation of the material in digestible sections not to overload the working memory (Sweller, van Merrienboer, & Paas, 1998; van der Meij, 2018) and to let learners have control over the flow of information (Mayer & Fiorella, 2018). In addition, compact videos are assumed to enhance engagement (Guo, et al., 2014). Therefore, the length of each video varied from 72 seconds to 158 seconds maximum per a video.

The first video was the introduction and aimed at brief presentation of the content of the six videos tutorials. Its goal was also to gain attention of learners and stimulate their interest in the training by explaining the benefits of mastering the to-be learned skills. The six video tutorials demonstrated an algorithm of the Word menu commands to complete a task.

All the six video tutorials had a similar order of presenting the material. The videos were produced in accordance with the cognitive theory of multimedia learning (Mayer, Heiser & Lonn, 2001) and the guidelines for the design of instructional videos for software training (van der Meij & van der Meij, 2013). The video animation was narrated based on the assumption, that verbal and visual information would be processed by different channels and that both channels would need to be used in a balanced way for effective learning. In each video, a to-be learned task procedure was demonstrated on the computer screen in a flawless manner by a model (see van der Meij & van der Meij, 2013). The model narrated the video to ensure association between the narrative and the video animation, which is believed to enhance learning (Mayer, 2008). Screenshots of the core video tutorial fragments with corresponding fragments of the narration are presented in Appendices G-L.

Each video was further segmented into subsections in accordance with sub-goals, such as "Explain and demonstrate acronym usage" and "Create an acronym replacement", aimed at motivating the learners (van der Meij, van der Meij, Voerman, & Duipmans, 2017). Each video started with a brief preview of a to-be-learned task demonstrating an ultimate goal and then presented the steps to complete the task on the MS Word interface. In the preview, the narrative included motivational sentences spoken in a conversational style by a model using personal pronouns "we, you", such as: "Instead of typing this over and over again in full, you can use acronym." Besides, because the focus audience was university students, for a greater motivation, the model mentioned that the students would be able to use the skills in writing their final thesis paper.

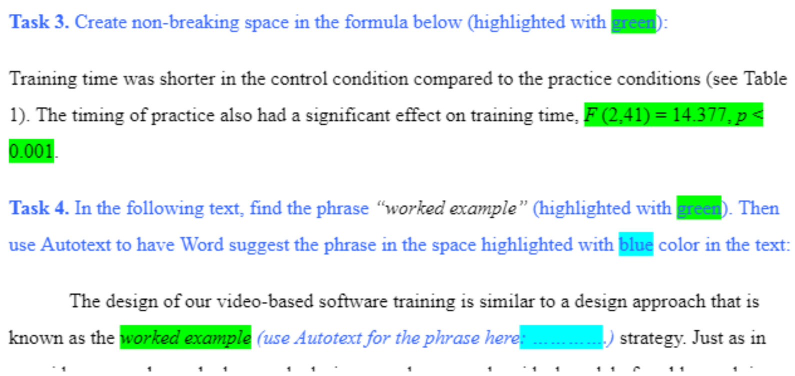
Procedural instructions explain how to have a task completed. Therefore, representing users' actions and the system's response (i.e., reaction) to the actions is key in design of procedural instructions (Steehouder & van der Meij, 2005). To effectively present the action-reaction information in the instructional videos, the guideline of action-reaction was adopted in design of the instructional videos (see van der Meij and Gellevij, 2004). The action part instructed a learner on what steps to do using imperative verbs, for instance: "you type demonstration-based training". The reaction part was a demonstration of what was expected to happen on the screen following execution of a task. In the end of each video, the result of the actions was demonstrated. Also, to draw attention of a learner to a specific item on the screen, zooming and signalling techniques were used (e.g., marking items with red line).

4.2.2. Word practice and test files

Five MS Word documents were created: one pre-training task performance test, one post-training task performance test, and one Word document per each of the three conditions for practicing the taught skills during training (see Figure 2). The Word documents included instructions on each task that were followed by a short Word text fragment used by participants to practice the learned skills. The Word documents for practice and testing were identical in the underlying structure but different in surface features and in the number of tasks to be executed. The practice document for each condition contained eleven Word-processing tasks to practice the six Word-processing skills taught in the instructional videos. Each Word-processing skill was practiced twice, except for the "creating a caption" task that was practiced only once due to the structure of the cumulative schedule. The tasks in the practice Word documents were presented in accordance with the orders prescribed by each condition (see Appenix B). The pre-training and the post-training task performance tests contained six tasks each, one task per each Word-processing skill taught. The Word files were shared with participants via respective Graasp spaces using the Google Drive tool.

Figure 2

A Fragment from a Word Task Practice Document


Task 3. Create non-breaking space in the formula below (highlighted with green):

Training time was shorter in the control condition compared to the practice conditions (see Table 1). The timing of practice also had a significant effect on training time, $F(2,41) = 14.377, p < 0.001$.

Task 4. In the following text, find the phrase *"worked example"* (highlighted with green). Then use Autotext to have Word suggest the phrase in the space highlighted with blue color in the text:

The design of our video-based software training is similar to a design approach that is known as the *worked example* *(use Autotext for the phrase here: ………….)* strategy. Just as in

4.2.3. Instructional booklets

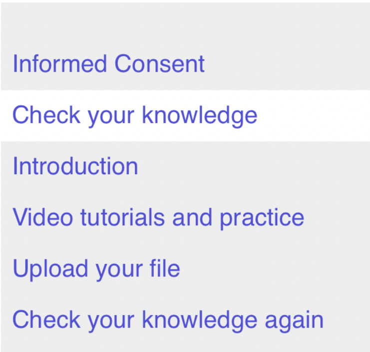
Instructional booklets were created in Word documents for each condition (i.e., a blocked, an interleaved, and a cumulative practice conditions). Each booklet guided participants through the experimental stages in accordance with a sequence prescribed by each condition (see Appendix B). The booklets included guiding instructions, as well as suggestions of a maximum time for completion of tasks and a rest-time between the stages to control duration of the experiments. Also, the booklet reminded of the possibility to replay videos during training to ensure presence of external feedback as suggested in the van der Meij and Maseland's (2021) study. The booklets served as a basis for construction of the learning environments in the Graasp online platform as described in section 4.2.4. Certain technical aspects, such as how to start playing videos and download and upload files through the platform, were shared in the booklets with some visual representations.

4.2.4. The learning environment

The learning environments were designed in the Graasp platform (see Appendix A for the web-links). The platform allows to create learning environments (i.e., spaces) presenting materials and instruments in an online mode. Learners could choose answers in multiple-choice question items, answer the Likert-scale questions by sliding a scale, play the videos, and download and upload the Word documents. Instructions and activities were organised in 6 sections that participants could navigate through by clicking options of the content menu on the left-hand side of the screen (see Figure 3).

Figure 3

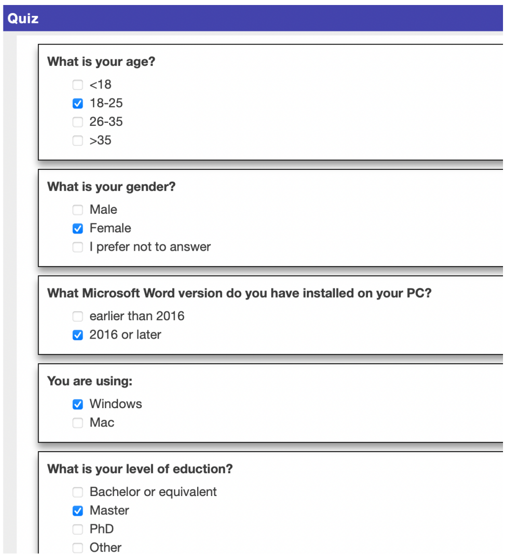
A Screenshot of the Graasp Space Menu



The Word documents for practicing tasks and testing the procedural knowledge were downloadable through the Google Drive web-links pasted in the respective sections of the spaces. Videos were embedded in the spaces using Youtube web-links. In the Graasp spaces, initially, participants were introduced into the theme of the research and asked to answer a few questions, such as age and a Word version being used for the experiment (see Figure 4).

Figure 4

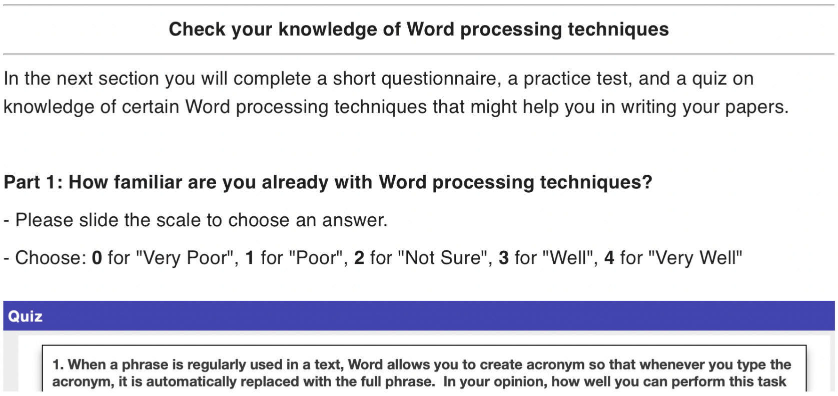
A Screenshot of a Fragment of the Initial Questionnaire from the Graasp Space



Before engaging in training, participants were asked to complete the self-efficacy questionnaire, as well as the procedural and the conceptual knowledge pretests (see Figure 5).

Figure 5

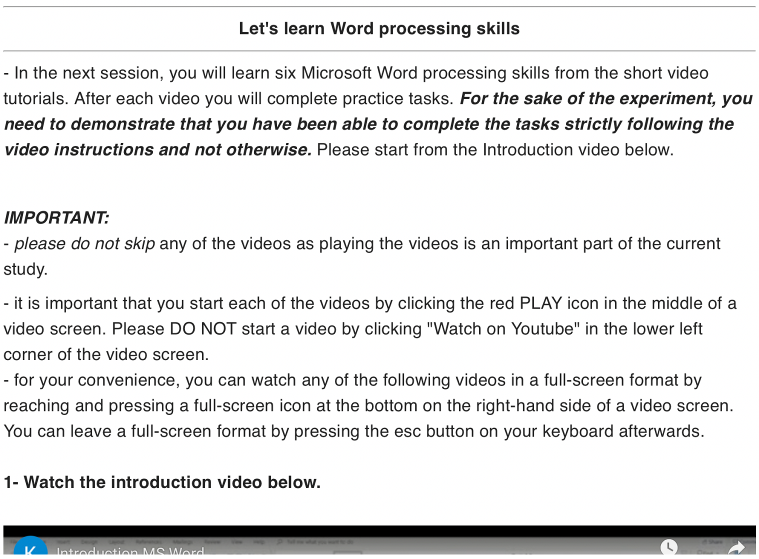
A Screenshot of the Pre-training Phase Fragment from the Graasp Space



Next, in the introduction section, participants were instructed to watch the introduction video, download a Word document corresponding to the condition and to proceed to the next section. The training content (i.e., the video tutorials and practice tasks) was delivered in the video tutorials and practice section, in accordance with the orders prescribed by the experimental conditions (see Appendix B and Figure 6).

Figure 6

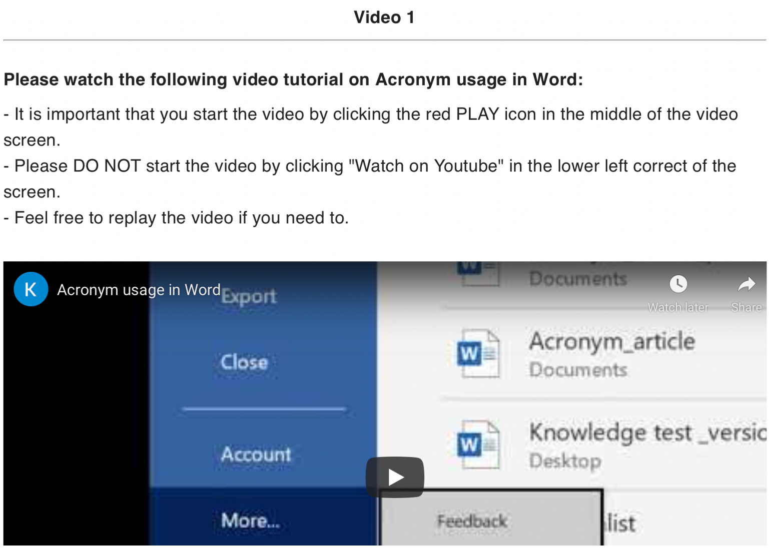
A Screenshot of the Introductory Instructions from the Graasp Space



Next, participants were invited to watch videos in accordance with the orders prescribed by the experimental conditions (see Appendix B and Figure 7). Instructions to each video included a reminder that the video may be replayed if needed.

Figure 7

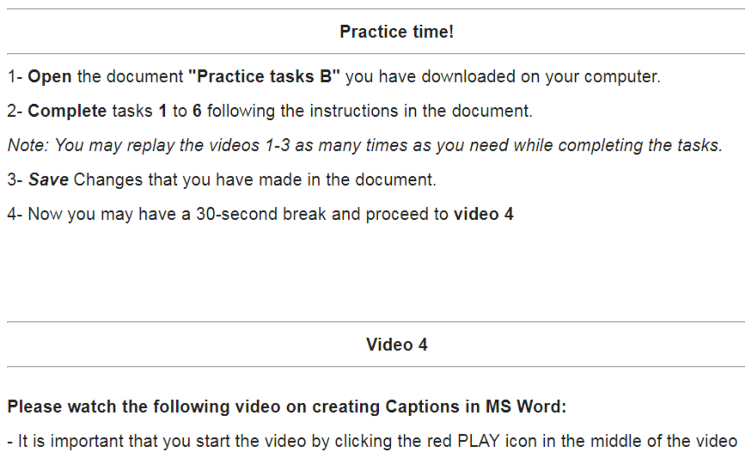
A Screenshot with the Video Tutorials and Practice Phase Fragment from the Graasp Space



After watching each video in the blocked condition or a series of instructional videos in the mixed conditions, participants were referred to the specific tasks in the Word practice document in accordance with the order prescribed by the experimental conditions (see Figure 8 and Appendix B). A one 30-minute break was suggested during training, between practice blocks in each condition. A 60-minute break was recommended before completing the questionnaires and tests after training.

Figure 8

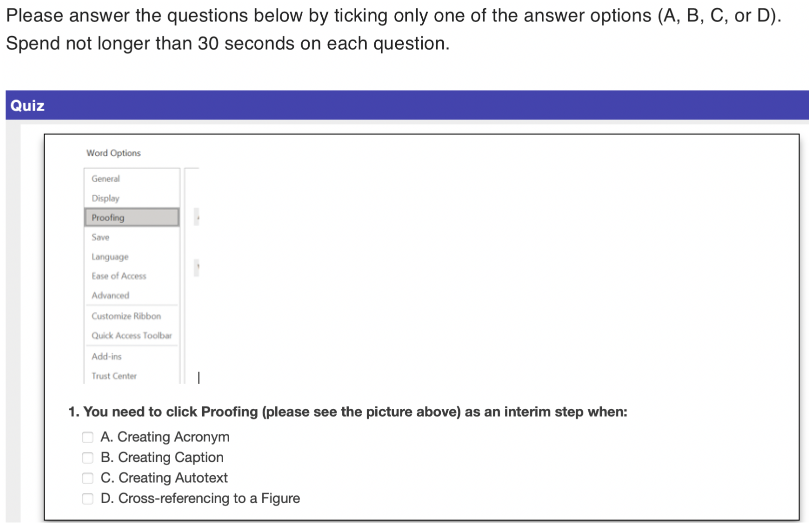
A Screenshot of the Practice Instructions from the Graasp Space



After watching the videos and practicing all the tasks in the video tutorial and practice phase, participants were asked to complete the self-efficacy and preference questionnaire, as well as the procedural and the conceptual knowledge posttests (see Figure 9). At this stage, participants were instructed not to replay the video tutorials any further and reminded that they should upload all the three Word documents with the completed tasks.

Figure 9

A screenshot with the Conceptual Knowledge Test Fragment from the Graasp Space



4.3. Instruments

4.3.1. Self-efficacy

Self-efficacy was measured with nine items. The items were developed in accordance with the Bandura's (2006) guide for constructing a self-efficacy scale and the self-efficacy scale developed by van der Meij (2018). Pre- and post-training self-efficacy items included identical questions. The questions asked how well a participant could use an appropriate algorithm to complete the Word-processing tasks as taught in the videos (see Appendix E). Answers were scored with a five-point Likert-scale and ranged from 0 (very poor) to 4 (very well).

The scale had a high level of internal consistency, as determined by a Cronbach's alpha of 0.818 and 0.916 for self-efficacy pre-training and post-training questionnaires, respectively.

4.3.2. Engagement

A view logging application was used in the Graasp spaces to track the number of seconds each video was played in total (i.e., total number of seconds a video has been played, including replays), unique playtime (i.e., the length of a video minus the length of the video parts not played at all, in seconds), and replay time (i.e., total number of seconds parts of a video have been replayed). If parts of a video were replayed, the total playtime for the video could exceed the length of the video. Unique playtime informs of coverage of the taught material and, If a video was played at least once in full, unique playtime for the video would be equal to the length of the video. If a video was played at least once but skipping some parts of the video, unique playtime for the video would be the difference between the length of the video and the total length of the parts of the video not played even once. Unique playtime could not exceed the length of a video. Replay time included total duration of a video parts that were replayed after initial viewing. Replay time could exceed the length of a video. While any of the described data could serve as a proxy for engagement, unique playtime and replay time were used as measures of engagement as they reflected the quality of viewing the videos. Engagement was calculated for all videos together.

4.3.3. Conceptual knowledge

There were two assessments of conceptual knowledge: pretest and posttest. Pretest and posttest included six identical multiple-choice questions on each of the six learned Word-processing skills. In the conceptual knowledge tests, a participant was asked to identify MS Word commands and algorithms required to complete the taught tasks. An example of an item is: "What is the correct sequence of commands for keeping the header (e.g., Header 1) and paragraph together (please see the picture above)?" Four answer options followed each item, with one answer option only being correct.

The order of answer options in the pre- and posttest differed but was the same between the conditions for each test. No answer or incorrect answer for multiple-choice questions was assigned 0 points. Correct answer was assigned 1 point. A maximum possible score was 6 per pre-test and post-test.

4.3.4. Procedural knowledge during and after training

There were three assessments of procedural knowledge: before training, during training, and after training. For the purpose of the pre- and post-training assessments, participants were asked to complete 6 Word-processing tasks, one per each taught skill, in a Word document that included instructional guidance and Word texts (see Appendix D). The instrument for assessment of procedural knowledge during training included 11 tasks for each condition (see Appendix C). The order of tasks was different between the three conditions. Design of interleaved and cumulative practice schedules was based upon compatibility of task completion algorithms corresponding to each of the to-be-learned skills from the perspective of creating "the desired difficulties" (see Appendices G, H, I, J, K, and L). 1 point was given for every correctly completed task. A task was accepted as completed if correct commands in Word menu were used. 0 points were given if a task was not completed. Thus, the maximum score for pre- and for post-training procedural knowledge was 6, and for the practice during training the maximum score was 11. Reliability analysis showed

Cronbach's alpha of 0.578 for pre-training task performance test, 0.780 for post-training task performance test, and 0.862 for practice test (i.e., task performance during training).

4.3.5. The learners' preference

There were three items in the learners' preference post-training questionnaire measuring the learners' preference of a practice schedule after training. The items were: "In your opinion, how well does this software training support your studying habits that usually help you in learning new skills?", "How well does this software training meet your expectations of convenient training?", "How well does this training meet your expectations of effective training?". Answers were scored with the five-point Likert scale and ranged from 0 (very poor) to 4 (very well). Cronbach's alpha was 0.843 indicating a high reliability of the instrument.

4.4. Procedure

The experiment was administered in an online mode via Graasp Platform (see Appendix A for the web-links). Participants completed the experiment in the spaces created for each condition in the Graasp platform from their individual locations at the time convenient for them. In all the three conditions, participants started from reading the introduction. In the introduction section participants were briefly informed of the nature of the experiment, invited to fill out the consent form and to give information, such as age, gender, the academic degree level being pursued, and a version of Word being used. For motivation purposes, it was mentioned that the videos would demonstrate Word-processing tasks needed for writing final theses. Further, it was explained that participants would watch short videos and practice tasks and that they should complete the tasks as taught in the video tutorials and not otherwise. Procedures and activities were executed in accordance with the specific architectures of the respective practice schedules as described in a more detail in the section 4.2.2. Total training time was approximately fifty-sixty minutes in each condition.

4.5. Data analysis

A Chi-squared test showed no difference between the conditions for gender, $\chi 2$ (2) = 1.1146; p = 0.564, and for age group, $\chi 2$ (4) = 1.005; p = 0.909. One-way ANOVA tests were run to compare means between the conditions. First, assumptions of normality and outliers were tested. Wherever the assumptions required for running ANOVA test were violated, the non-parametric Kruskal-Wallis test was conducted. Testing was with alpha set at 0.05. Cohen's (1988) d-statistic was used to compute effect sizes. The effect size is considered as small for d = 0.20, medium for d = 0.50, and as large for d = 0.80.

5. Results

5.1. The effect of a schedule of practice on self-efficacy

Table 4 shows that total self-efficacy scores before training were relatively neutral with the lowest score found in the cumulative practice condition. After training, the scores were substantially closer to the maximum possible total score value. The lowest post-training score was found in the interleaved practice condition. No significant difference was found between self-efficacy gain scores across conditions, $F(2,45) = 1.09$, $p = 0.346$, $\eta^2 = 0.046$, indicating that the extent of self-efficacy development was not influenced by a practice schedule. Eta-squared value suggested that only 4.6% of the variability in the self-efficacy change scores could be attributed to the type of practice schedule applied. The effect-size, $d = 0.482$, was small to medium.

Table 4

Means[a] (and standard deviation) for self-efficacy per condition and test moment

Condition	Pre-training		Post-training	
	M	SD	M	SD
Blocked (*n* = 16)	52.61	22.08	92.03	10.58
Interleaved (*n* = 15)[b]	47.03	20.11	79.64	25.92
Cumulative (*n* = 17)	43.47	19.11	88.56	13.03
Overall (n = 48)[b]	*47.92*	*20.39*	*86.92*	*17.89*

Note. [a] means of total scores per test based on the scale values ranging from 0 to 4 per item, with higher values meaning a more positive rating; in percentages of a maximum possible score for the test; [b] n after removal of an outlier.

5.2. The effect of a schedule of practice on engagement

The findings on engagement are presented in Table 5. The unique play time was comparable across the three conditions with a slightly higher value in the interleaved condition. Total playtime was relatively uniform across the conditions and comparable with the total duration of all six videos (674 seconds). In addition, the replay time was higher in the interleaved condition (9.14%) suggesting that participants were in a greater need to watch some parts of the videos more than once than in the two other conditions.

In addition, in terms of unique playtime accepted as a proxy for engagement, there was no significant difference between the groups, $\chi2(2) = 3.174$, $p = 0.205$. This finding suggests that schedule of practice did not have an influence on coverage of the taught material. That is, in all the three conditions, participants were equally engaged in training. As video replay possibility was rarely used and only by a few participants, the data was not further analysed.

Table 4

Means (and standard deviation) for engagement per condition

	Engagement					
	Total play[a]		Unique play[b]		Replay[c]	
Condition	M	SD	M	SD	M	SD
Blocked *(n = 16)*	628.69	292.17	74.06	31.57	2.65	6.18
Interleaved *(n = 16)*	739.63	314.35	82.22	26.51	9.14	20.41
Cumulative *(n = 17)*	641.06	333.77	74.59	36.87	1.10	2.88
Total (n = 49)	669.20	311.72	75.82	31.41	4.48	13.68

Note. [a] in seconds; [b] means and standard deviations in percentage of total duration of all six video tutorials; [c] means and standard deviations in percentage of the mean total video playtime per condition for all video tutorials.

5.3. The effect of a schedule of practice on learning

Task performance during training. There was no difference in the practice scores between the conditions with median practice scores being not significantly different between the groups, $\chi 2(2) = 0.138$, $p = 0.933$. Moreover, median practice scores for practice tasks completed during training were of the same value (*Mdn* = 11.00) for all the three conditions. In other words, there was no impact of a practice schedule on scores for task performance during training.

Table 5

Means[a] (and standard deviation) for task performance during training per condition

Condition	Training	
	M	SD
Blocked (*n* = 16)	92.64	9.52
Interleaved (*n* = 16)	90.36	16.42
Cumulative (*n* = 16)[b]	88.64	20.32
Overall (n = 48)[b]	*90.55*	*15.80*

Note: [a] means and standard deviations in percentage of the maximum possible score of 11 per test; [b] n after removal of an outlier.

Procedural knowledge. Results for procedural knowledge are presented in table 6. Notably, prior knowledge scores were the lowest in the interleaved practice condition but post-training task performance scores were comparable with those in the two other conditions. The mean rank of the procedural knowledge gain scores was not significantly different between the groups, $\chi 2(2) = 3.014$, $p = 0.222$, indicating that there was no impact of practice schedule on the procedural knowledge gain.

Table 6

Means[a] (and standard deviation) for procedural knowledge per condition

Condition	Pre-training		Post-training	
	M	SD	M	SD
Blocked (*n* = 16)	25.00	22.77	89.67	15.95
Interleaved (*n* = 16)	14.67	14.75	87.50	22.73
Cumulative (*n* = 17)	24.50	27.72	80.33	26.50
Overall (n = 49)	*21.50*	*22.57*	*85.67*	*21.52*

Note. [a] means and standard deviations in percentage of the maximum possible score of 6 per test.

Conceptual knowledge. Results presented in table 7 show that prior conceptual knowledge scores were uniform across the conditions. So were post-training conceptual knowledge scores. There was no statistically significant difference in conceptual knowledge gain from pretest to posttest between the conditions. Hence, conceptual knowledge gain was not influenced by practice schedule, $F(2,46) = 1.09$, $p = 0.028$, $\eta^2 = 0.973$.

Table 7

Means[a] (and standard deviation) for conceptual knowledge per condition

Condition	Pre-training		Post-training	
	M	SD	M	SD
Blocked (*n* = 16)	42.67	26.50	80.17	15.18
Interleaved (*n* = 16)	43.67	19.12	81.33	11.98
Cumulative (*n* = 17)	42.83	22.92	81.33	17.57
Overall (n = 49)	*42.83*	*22.57*	*81.00*	*14.83*

Note. [a] means and standard deviations in percentage of the maximum possible score of 6 per test

5.4. The effect of a schedule of practice on preference

Results for preference of practice schedule by participants presented in table 8 show that preference scores were slightly lower in the interleaved condition than in the two other conditions. However, the median scores were not significantly different between the groups: blocked (*Mdn* = 11.50), interleaved (*Mdn* = 11.00), and cumulative (*Mdn* = 11.50), $\chi 2(2) = 0.375$, p = 0.829. The findings indicate that there was no difference in preference of a practice schedule among the participants between the three practice conditions.

Table 8

Means[a] (and standard deviation) for preference per condition

Condition	Preference	
	M	SD
Blocked (*n* = 16)	88.58	14.55
Interleaved (*n* = 16)	82.83	20.97
Cumulative (*n* = 17)	96.27	13.51
Overall (n = 49)	*86.58*	*16.48*

Note. [a] means and standard deviations in percentage of the maximum possible score of 12 for 3 items, based on the Likert scale ranging from 0 to 4, with higher value meaning a more positive rating.

6. Discussion and conclusion

The current study compared the effects of blocked and mixed schedules of practice on self-efficacy, engagement, learning, and preference in software training delivered through video tutorials. For the purpose of comparison with the effect of a blocked practice, the current study aimed at triggering the CI-effect with the mixed practice schedules. In addition, the boundary conditions for the CI-effect suggested in previous research were facilitated in this study.

The first research question (RQ1) concerned the effects of practice schedules on learners' self-efficacy. It was expected that the combination of blocking and interleaving inherent in a cumulative practice would have a greater influence on a mastery experience (see Meij & Maseland, 2021). In its turn, the latter, in accordance with Bandura (1997), would be crucial for positive changes in self-efficacy. However, just like in the recent study of van der Meij and Maseland (2021), no influence of a practice schedule on self-efficacy development was established.

In the van der Meij and Maseland's (2021) study, the outcome was partially explained with the high rate of completion of the tasks by all participants during practice, which could have equally affected confidence of learners in both, the blocked and interleaved conditions. Similarly, in the current study, a low number of zero scores for practice tasks could be an indicator of a high rate of completion, too. Another explanation for the non-significant test result could be pairing practice with the instructional videos whose architecture was aiming not only at serving "as a job aid" but at "enhancing learning" (van der Meij, 2014, p. 112). Possibly, observation of task completion through the effective video tutorials could have wiped out differences in the influence of practice schedules on self-efficacy.

Next, in the interleaved condition, the spread of the post-training self-efficacy scores was greater than in the two other conditions. The higher spread indicates that the learners who followed training based on the interleaved practice were less homogeneous in judgement of their self-efficacy after training. The study of van der Meij and Maseland (2021), had a similar finding but the spread was less striking. Perhaps, an interleaved practice was a cause of greater confusion among learners, which may have affected their confidence in their abilities. For comparison, the consistency in presentation of tasks in training based on a cumulative practice could be a reason for greater homogeneity of learners from the self-efficacy perspective.

The second research question (RQ2) concerned the effect of practice schedules on learners' engagement in training. Contrary to the expectations, in the interleaved practice condition, engagement in terms of unique video playtime reflecting video coverage was not significantly different than in the other two conditions. First, the finding that the coverage did not reach 100% in either of the conditions could be an indication of an effective segmentation of each video. Perhaps, the structure of the videos allowed participants to get an easy orientation and skip the introductory parts of the videos as soon as they felt informed well enough of a to-be-learned skill. At the same time, the average coverage was substantially higher than the length of introductory parts of the videos. Presumably, participants needed to cover the instructional parts of the videos in full, regardless of a practice schedule, as the videos taught advanced Word-processing skills.

The finding that participants did not take advantage of the video replay possibility could be related to the retrieval strength that is believed to be crucial for immediate performance during practice (see Soderstorm & Bjork, 2015). More specifically, in software training, the capacity to retrieve information (i.e., retrieval strength) from the working memory might have been equally enhanced in all the three conditions by the aiding properties of the software interface menu. As a result, the external feedback in the form of the video replay was not demanded in either of the conditions, including the interleaved practice condition. In addition, the content and the architecture of the videos might have been effective enough to avoid replaying a video. Also, it would be fair to mention that, in all the conditions, the mean value of total playtime was close to the total length of all videos. Notably, with interleaved practice, it exceeded the total length of the videos, which is in agreement with the finding of the highest replay time rate in the condition. Whereas, unlike unique playtime and replay time, this finding cannot serve as an evidence of engagement quality, it reflects the level of dedication of learners to training. One way or another, as participants hardly replayed the videos, this study did not give a reason to judge to what extent the feedback frequency could affect learning (see Schmidt and Bjork, 1992) in the video-based software training. Overall, in software training, the “desired difficulties” (see Bjork & Bjork, 2011) inherent in the mixed practice schedules did not result in a greater time investment or in a greater knowledge gain. Therefore, this study failed to facilitate a ground for supporting the finding that a greater time investment into a study would result in better learning (see Kuh, 2003).

The next research question (RQ3) concerned the effects of practice schedules on learning, namely, the effects on procedural and conceptual knowledge gain after training, as well as on task performance during training. In accordance with the latest research, with the blocked schedule, task performance score was consistently higher during practice, immediate and delayed tests, although the difference was not significant (van der Meij & Maseland, 2021). Likewise, in the current study, there was no influence of a practice schedule either on task performance scores during training or on a conceptual and procedural knowledge gain after training. In brief, in the current research, with video-based software training, presence of the CI-effect was confirmed neither with a procedural nor with a conceptual knowledge gain. Moreover, task performance during training was not worse with the mixed schedules of practice than with blocked practice. This contradicts to the robust findings in other domains (e.g., Rohrer et al., 2020; Taylor & Rohrer, 2010) where blocked practice repeatedly outperformed interleaved practice at the acquisition stage. Nevertheless, in line with the conclusions of van der Meij and Maseland (2021), no difference was found in the effect of blocked and mixed schedules of practice in the video-based software training either on knowledge gain or on task performance at the knowledge acquisition stage.

Van der Meij and Maseland (2021) partially attributed the inconsistency of the posttest findings to the lack of practice trials in their study where each task was practiced with one trial only. The explanation was based upon examples from studies in motor task and mathematics learning where presence of the CI-effect was noted with multiple practice trials per skill (e.g., Rohrer et al., 2020; Taylor & Rohrer, 2010). Likewise, practicing the taught skills twice during training in the current study might have not been still sufficient enough to increase the testing effect (see van der Meij & Maseland (2021). On the other hand, more than two practice trials per skill would inevitably lead to longer training and, consequently, to a lower time-efficiency of mixed schedules of practice. Therefore, in the view of replication of the van der Meij and Maseland's (2021) findings, from the practical point of view, complicating of software training with mixed schedules of practice at the cost of time-efficiency seems to be unnecessary. In the future research, however, it would be interesting to investigate whether more practice could result in a speeded-up task performance. Indeed, given the insignificant difference in procedural knowledge development, choice of a practice schedule could be made in favour of the schedule ensuring a faster task execution with an extra practice.

Another explanation van der Meij and Maseland (2021) gave for the lack of support of the well-established findings was that participants did not have a possibility of an external feedback in their study. Van der Meij and Maseland (2021) argue that replaying the videos during practice would facilitate a feedback on whether a learner's action has been correct. For comparison, an external feedback was present during training in some studies that resulted in the CI-effect in other domains (e.g., Helsdingen et al., 2011; Rohrer, et al., 2020). However, in the current study, despite the possibility to replay videos during practice, the CI-effect was established with none of the two mixed schedules. At the same time, it is fair to mention that far not all participants took advantage of the replay possibility. On one hand, the low replay rate might be an evidence of the relatively good quality and a clear content of the videos. In addition, as the experiment took approximately one hour, participants might have merely avoided spending extra-time on replaying videos, regardless of the skill mastery level. On the other hand, software itself immediately gives feedback signalling of wrong choices of commands at each step (van der Meij & Maseland, 2021). Therefore, the need in an external feedback (i.e., the need in video replay) may be not vital enough with video-based software training.

Nevertheless, a wider spread of scores for task performance during training and the post-training procedural knowledge test in both of the mixed practice conditions is a notable finding. It points out that, in these two groups, learners were less homogeneous in their task performance abilities. To compare, the van der Meij and Maseland's (2021) study shows a greater spread of procedural knowledge scores for the immediate retention test in the blocked practice condition, but a relatively equal spread of task performance scores in the blocked and interleaved conditions. The finding cannot be explained by the assumed synergy of instructional videos and the improved interface menu because it was true for the van der Meij and Maseland's (2021) study, too. One possible explanation could be that participants in the latter study were children and there was no external feedback possibility. Presumably, primary school students would use video replay more intensively than adults, which could result in more balanced procedural test results. Similarly, availability of external feedback with children could lead to a greater homogeneity of learners in their task performance during training.

To conclude, with adult learners, complicating practice schedules by mixing the tasks did not serve its initial purpose. Indeed, given that all three schedules are equally effective, a practice schedule, which facilitates more homogeneity of subjects from the perspective of the procedural task execution abilities, is more advantageous. Consequently, a blocked practice schedule seems preferrable in the video-based software training, particularly, for adult learners.

Overall, in the current study, the boundary conditions expected to contribute to the CI-effect failed to secure the success of the mixed practice schedules in learning. Therefore, comparing the conditions of the current study with previous research that supported presence of the CI-effect may shed the light on the issue and point to the directions for future research (see Table 9).

Firstly, previous research reveals that effectiveness of mixed schedules of practice in terms of retention depends on the structure of the categories learned. If the difference between categories is subtle, then there is a potential for the CI-effect presence with mixed practice (see Firth, 2018; Kornell and Bjork, 2008; Rohrer, Dedrick, & Burgess, 2014). In the current study, however, the task completion algorithms for the learned skills were rather highly dissimilar. For instance, it is unlikely to confuse the combination of commands to create a caption with the action required to create a non-breaking space (see Appendices I and K). Moreover, with the video-based software training, the freedom of manipulations with discriminability of task completion routes is restricted due to the fixed interface menu. Therefore, the main prerequisite condition for the presence of the CI-effect is hardly achievable.

Table 9

The checklist of critical criteria for presence of the CI-effect with mixed practice schedules

Criteria	Compliance with the criteria		
	Yes	No	N/A
Is learning deductive (rather than inductive)?	●		
Does each to-be learned skill require an isolated and independent knowledge?	●		
Are the taught skills or categories or pieces of knowledge dissimilar?	●		
Is practice schedule to be chosen by a learner?		●	
Does a testing method match a practice schedule and the tested knowledge?		●	
Are the to-be-learned pieces of knowledge/categories highly discriminable?	●		
Is the to-be-learned material novel for learners?			●

Next, in previous research, a blocked schedule showed greater retention than mixed practice when a testing method was aimed at testing isolated, independent and dissimilar pieces of knowledge (see Carvalho and Goldstone, 2020). The latter finding is a derivative of the discriminative contrast hypothesis. In the current study, each word processing skill assumed rather an independent task completion algorithm. Following the Carvalho and Goldstone's (2020) logic, with mixed practice schedules in video-based software training, multiple-choice tests might have had a greater positive effect in learning less discriminable word-processing algorithms. This implies a selective approach to mixing tasks. For instance, if a task is to keep a header and a paragraph together in a text, perhaps, there may be a confusion between the sub-commands, such as "keep with next", "keep lines together", and "page break before" (see Appendix L). It is important to remember, however, that a decision on confusability of tasks is rather subjective. Therefore, even if the Carvalho and Goldstone's (2020) finding is supported, this may be a result of interaction of other multiple possible causes. In brief, the current finding reveals that, for practitioners, there is not sufficient evidence yet to consider a form of testing as the decision criteria in choosing between the practice schedules for training.

Third, mixed schedules of practice seem to have been effective with inductive learning (see Carvalho & Goldstone, 2014; Kang & Pashler, 2012; Kornell & Bjork, 2008). Despite the deductive nature of learning in the current study (i.e., rules were explained in a detail in the videos prior to practicing the tasks; see Carvalho & Goldstone, 2014), the effect of the practice schedules was the same for all conditions. Therefore, a theoretical implication would be that, with video-based software training, a type of learning might not play a decisive role in development of knowledge as it does in other domains. On one hand, future investigations of the effect might be conducted designing a more inductive way of learning. On the other hand, this task may turn to be rather complicated in learning software skills through videos. Objectively, facilitation of induction seems to be hardly achievable while explicitly demonstrating the Word-processing procedures through videos. Therefore, with video-based software training, practitioners are free to choose the least complicated and time-consuming way of practice. Indeed, the current study reveals that blocked practice brings to the same knowledge outcome as schedules that are more complex.

Forth, blocked and cumulative schedules were suggested as more effective with novice learners (see Carvalho & Goldstone, 2014; Mayfield & Chase, 2002). In the current study, prior knowledge of participants in the three conditions was approximately the same with a slightly higher mean value in the blocked condition. Therefore, it could be assumed, that the material was of equal novelty for all participants in all three conditions. Consequently, the level of novelty of the material for participants in the blocked or cumulative conditions was not responsible for the practice schedules being as effective as the interleaved schedule.

Finally, in earlier studies, a better retention was achieved if a practice schedule was deliberately chosen by learners (see Carvalho et al., 2016). In the current study, participants were randomly assigned to conditions and were not made aware of what practice schedule they would be following. On one hand, the finding does not contradict to the conclusions of Carvalho et al. (2016). On the other hand, the capacity of learners in assessing their learning efforts and the resulting learning outcomes associated with a particular learning mode is disputable. Thus, a deliberate choice of a practice schedule is not necessarily well-thought and can be coincidental. Overall, software developers can deliver video-based software training with the least complicated practice schedule.

The final research question (RQ4) concerned the effect of training with different practice arrangements on learners' preference. Contrary to the expectations, participants in all the three condition equally preferred a particular training mode they followed. In addition, the data shows that training appreciation scores were far above the neutral score in all the three conditions. However, in other domains, participants favoured training with blocked practice although it did not necessarily result in a better retention. The main explanation for the equal preference in this study remains to be the improved software usability that made participants insensitive to "the desirable difficulties" (see Bjork & Bjork, 2011) intentionally created in the mixed schedules. In other words, the immediate feedback given by the menu could have helped participants in the mixed conditions gain the same feeling of ease and fluency as in training with the blocked practice.

Nevertheless, in the interleaved practice condition, learners were less homogeneous in their preference of the training method, whereas, in the other two conditions, learners' preference was less diverse. The greater unity of learners in the perception of training with a blocked and cumulative practice schedule as more convenient and effective is in agreement with the assumption of the perceived sense of ease and fluency associated with studying based on a blocked method (see Kornell and Bjork, 2008). Because a cumulative practice has blocking properties, the assumption is fair for this type of practice as well. However, it is worth of reminding that the judgement of a practice schedule is subjective (see Kornell and Bjork, 2008) and can depend on multiple factors.

Mentioning limitations of the current study, first, the lengthy experiment did not allow gaining a greater number of quality participants within the given timeline. Secondly, because of the need to control the duration of the experiment, the number of to-be-learned MS Word-processing skills was restricted. In its turn, the small variety of taught skills limited the possibility for manipulations with discriminability of the task completion routes in the mixed schedules. Third, the tests were conducted during a single session only that restricted the effect of forgetting, which was suggested to be vital with mixed practice schedules (see Soderstorm & Bjork, 2015).

To conclude, under the Covid-19 circumstances, online learning is literally the only way to access study material without an interruption from any location. At the times when learning from observation of real models has become insecure, learning from video-tutorials seems to be the most realistic solution. In addition, it is a relatively easy-to-implement solution with software training. Therefore, contribution to investigation of potentially effective studying arrangements cannot be underestimated. With software training, a right combination of justified practice schedules with instructional videos constructed in accordance with the cognitive theories could be the unique solution for effective learning. The main input of this study is that software training developers can be flexible and more creative designing video-based training than they currently are, subject to resource restrictions. As for academia, the findings of this study are a favourable ground for advocating practice-inclusive video-based teaching in educational establishments and for general public. Perhaps, the practice-inclusive approach could be enriched further with additional techniques, such as collaborative learning from videos, learning that aims at a faster task performance, interactive learning or even a discovery-based learning (e.g., with a younger audience). The latter suggestions could be a direction for future research.

7. References

Bandura, A. (1986). Social foundations of thought and action. A social learning theory. Englewood Cliffs, NJ: Prentice-Hall.

Battig, WF. (1966). Facilitation and interference. In: Acquisition of Skill, edited by Bilodeau E. New York: Academic, 215-244.

Birnbaum, M., Kornell, N., Bjork, E., and Bjork, R. (2013). Why interleaving enhances inductive learning: the roles of discrimination and retrieval. *Mem. Cogn.*, *41*, 392-402. http://doi.org/10.3758/s13421-012-0272-7

Bjork, E. L., & Bjork, R. A. (2011). Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. In M. A. Gernsbacher, R. W. Pew, L. M. Hough, J. R. Pomerantz (Eds.) & FABBS Foundation, Psychology and the real world: Essays illustrating fundamental contributions to society (p. 56–64). Worth Publishers.

Bjork, R. & Bjork, E. (2019). Forgetting as the friend of learning: implications for teaching and self- regulated learning. *Adv Physiol Educ, 43*, 164–167. doi:10.1152/advan.00001.2019

Bjork, R., & Bjork, E. (2020). Desirable difficulties in theory and practice. *Journal of Applied research in Memory and Cognition, 9* (4), 475-479.

Blandin, Y., Proteau, L., & Alain, C. (1994). On the cognitive processes underlying contextual interference and observational learning. *Journal of Motor Behavior, 26*(1), 18-26. http://doi.org/10.1080/00222895.1994.9941657

Bouzid N., & Crawshaw, C. (1987). Massed versus distributed wordprocessor training. *Applied Ergonomics, 18*(3)*, 220-222.*

Brady, F. (1998). A theoretical and empirical review of the contextual interference effect and the learning of motor skills. QUEST, 50, 266–293.

Brunmair, M., & Richter, T. (2019). Similarity matters: a meta-analysis of interleaved learning and its moderators. *Psychological Bulletin, 145*(11), 1029-1052.

Carpenter, S., Cepeda, N., Rohrer, D., Kang, S., & Pashler, H. (2012). Using spacing to enhance diverse forms of learning: Review of recent research and implications for instruction. *Educational Psychology Review*, *24*, 369-378.

Carpenter, S., & Mueller, F. (2013). The effects of interleaving versus blocking on foreign language pronunciation learning. *Memory & Cognition*, *41*, 671-682. http://doi.org/10.3758/s13421-012-0291-4

Carvalho, P., & Goldstone, R. (2014), Effects of interleaved and blocked study on delayed test of category learning generalization. *Frontiers in Psychology, 5*. http://doi.org: 10.3389/fpsyg.2014.00936

Carvalho P., & Goldstone R. (2014). Putting category learning in order: category structure and temporal arrangement affect the benefit of interleaved over blocked study. *Mem Cognit., 42*(3), 481-95. http://doi.org/10.3758/s13421-013-0371-0

Carvalho P., Braithwaite D., de Leeuw J., Motz B., & Goldstone R. (2016). An in vivo study of self-regulated study sequencing in introductory psychology courses. *PLoS ONE* 11(3): e0152115. doi:10.1371/journal.pone.0152115

Carvalho, & Goldstone, (2020). The most efficient sequence of study depends on the type of test. *Applied Cognitive Psychology, 35*(1), 82-97. https://doi.org/10.1002/acp.3740

Chen, O., Paas, F., & Sweller, J. (2021). Spacing and interleaving effects require distinct theoretical bases: a systematic review testing the cognitive load and discriminative-contrast hypotheses. *Educational Psychology Review*. https://doi.org/10.1007/s10648-021-09613-w

Cohen, J. (1988). Statistical power analysis for the behavioural sciences 92ND ED.). Hillsdale, NJ: Erlbaum.

Driskell, J., Willis, R., & Cooper, C. (1992). Effect of overlearning on retention. *Journal of Applied Psychology 77*(5), 615-622.

Dunlosky J, Rawson K., Marsh E., Nathan M., & Willingham D. (2013). Improving students' learning with effective learning techniques: promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest*, 14(1), 4-58. doi:10.1177/1529100612453266

Fiorella, L. & Mayer, R. (2018). What works and doesn't work with instructional video. *Computers in Human Behavior, 89*, 465-470. https://doi.org/10.1016/j.chb.2018.07.015.

Firth, J. (2018). The application of spacing and interleaving approaches in the classroom*.* *Journal of the Chartered College of Teaching*, 2, 23-26.

Firth, J., Rivers., I., & Boyle, J. (2021). A systematic review of interleaving as a concept learning strategy. *Review of Education*. http://doi.org/10.1002/rev3.3266

Foster, N., Mueller, M., Was, C., Rawson, K., & Dunlosky, J. (2019). Why does interleaving improve math learning? The contributions of discriminative contrast and distributed practice. *Mem Cogn, 47*, 1088–1101. https://doi.org/10.3758/s13421-019-00918-4

Grossman, R., Salas, E., Pavlas, D., & Rosen, M.A. (2013). Using instructional features to enhance demonstration-based training in management education. *Academy of Management Learning and Education, 12*(2), 219–243.

Guo, P., Kim, J., & Rubin, R. (2014). *How video production affects student engagement: An empirical study of MOOC videos*. In Proceedings of the first ACM conference on Learning@ scale conference, 41-50.

Helsdingen, A., van Gog, T., & van Merriënboer J. (2011). The effects of practice schedule on learning a complex judgement task. *Learning and Instruction, 21*(1), 126-136. http://doi.org/10.1016/j.learninstruc.2009.12.00

Higgins, E., & Ross, B. (2011). Comparisons in category learning: how best to compare for what. *Cognitive Science, 33*.

Hollender, N., Hofmann, C., Deneke, M., & Schmitz, B. (2010). Integrating cognitive load theory and concepts of human–computer interaction. *Computers in Human Behavior*, *26*(6), 1278-1288. https://doi.org/10.1016/j.chb.2010.05.031

Huang, H-W., Wu Ch-W., & Chen N-Sh. (2012). The effectiveness of using procedural scaffoldings in paper-plus-smart phone collaborative learning process. *Computers & Education*, *59*, 250-259. http://doi.org:10.1016/j.compedu.2012.01.015

Hughes, C., & Lee, J-Y. (2019). Effective approaches for scheduling and formatting practice: distributed, cumulative, and interleaved practice. *Teaching exceptional chi*ldren, *51*(6), 411-423. http://doi.org/10.1177/0040059919847194

Kang, S., & Pashler, H. (2012). Learning painting styles: spacing is advantageous when it promotes discriminative contrast applied cognitive psychology. *Appl. Cognit. Psychol.*, *26*, 97–103. http://doi.org/10.1002/acp.1801

Kornell, N., & Bjork, R. A. (2008). Learning concepts and categories: Is spacing the "enemy of induction"? *Psychological Science, 19*, 585–592. http://doi/10.1111/j.1467-9280.2008.02127.x

Lee, T., Swanson, L., & Hall, A. (1991). What is repeated in a repetition? Effects of practice conditions on motor skill acquisition. *Physical Therapy, 71*(2).

Leppink, J., Paas, F., van Gog, T., van der Vleuten, C., & van Merrienboer. J. (2014). Effects of pairs of problems and examples on task performance and different types of cognitive load. *Learning & Instruction*, *30*, 32-42. http://doi.org/10.1016/j.learninstruc.2013.12.001

Mayer, R.E., Heiser, J. & Lonn, S. (2001). Cognitive constraints on multimedia learning: When presenting more material results in less understanding. *Journal of Educational Psychology, 93*(1), 187-198.

Mayfield, K. H., & Chase, P. N. (2002). The effects of cumulative practice on mathematics problem solving. *Journal of Applied Behavior Analysis*, 35, 105–123. http://doi.org/10.1901/jaba.2002.35-105

Newell, A., & Rosenbloom, P. (1993). Mechanisms of skill acquisition and the law of practice. The Soar papers (vol. 1): research on integrated intelligence. MIT Press, Cambridge, MA, USA, 81–135.

Rau M.A., Aleven V., Rummel N. (2010). Blocked versus interleaved practice with multiple representations in an intelligent tutoring system for fractions. *In: Aleven V., Kay J., Mostow J. (eds) Intelligent Tutoring Systems. ITS 2010. Lecture Notes in Computer Science, vol 6094. Springer, Berlin, Heidelberg*. http://doi.org/10.1007/978-3-642-13388-6_45

Rogers, Y. (2004). New theoretical approaches for human-computer interaction. *Annual Review of Information Science and Technology*, *38*. http://doi.org/10.1002/aris.1440380103

Rohrer, D., Taylor, K., Pashler, H., Wixted, J. T., & Cepeda, N. J. (2005). The effect of overlearning on long-term retention. *Applied Cognitive Psychology*, *19*, 361-374.

Rohrer, D., & Taylor, K. (2007). The shuffling of mathematics problems improves learning. *Instructional Science*, *35*, 481–498.

Rohrer, D. (2009). The effects of spacing and mixing practice problems. *Journal for Research in Mathematics Education, 40* (1), 4-17.

Rohrer, D., Dedrick, R., & Burgess, K. (2014). The benefit of interleaved mathematics practice is not limited to superficially similar kinds of problems. *Psychonomic Bulletin & Review, 21*(5). http://doi.org/10.3785/s13423-014-0588-3

Rohrer, D., Dedrick, R., & Stershic, S. (2014). Interleaved practice improves mathematics learning. *Journal of educational Psychology*. http://dx.doi.org/10.1037/edu0000001

Rohrer, D., Dedrick, R., and Hartwig, M. (2020). The Scarcity of Interleaved Practice in Mathematics Textbooks. *Educational Psychology Review, 32*, 873–883. http://doi.org/10.1007/s10648-020-09516-2

Rohrer, D. & Hartwig, M. (2020). Unanswered questions about spaced interleaved mathematics practice. *Journal of Applied Research in Memory and Cognition 9*(4), 433-438. http://doi.org/10.1016/j.jarmac.2020.06.008

Shea, J. B., & Morgan, R. L. (1979). Contextual interference effects on the acquisition, retention, and transfer of a motor skill. *Journal of Experimental Psychology: Human Learning and*

Memory, 5(2), 179–187. http://doi.org/10.1037/0278-7393.5.2.179

Shea, Ch., Wright, D., Wulf, G., & Whitacre, C. (2000). Physical and observational practice afford unique learning opportunities. *Journal of Motor Behavior, 32*(1), 27-36. http://doi.org/10.1080/00222890009601357

Schmidt, R. A., & Bjork, R. A. (1992). New conceptualizations of practice: Common principles in three paradigms suggest new concepts for training. *Psychological Science, 3*, 207–217.

Soderstrom, N.C., & Bjork, R. (2015). Learning versus performance: an integrative review. *Perspectives on Psychological Science, 10*, 176 - 199. http://doi.org/ 10.1177/1745691615569000

Steehouder, M. F., & van der Meij, H. (2005). Designing and evaluating procedural instructions with the four components model. In *2005 IEEE International Professional Communication Conference Proceedings* (pp. 797-801). IEEE. http://doi.org/10.1109/IPCC.2005.1494254

Sundar S., & Nass C. (2000). Source orientation in human-computer interaction: programmer, networker, or independent social actor. *Communication Research, 27*(6), 683-703. http://doi.org/10.1177/009365000027006001

Suzuki, Y., Yokosawa, S., & Aline, D. (2020). The role of working memory in blocked and interleaved grammar practice: Proceduralization of L2 syntax. *Language Teaching Research*. https://doi.org/10.1177/1362168820913985

Sweller, J., van Merrienboer, J., & Paas, F. (1998). Cognitive architecture and instructional design. *Educational Psychology Review*, *10***,** 251–296. https://doi.org/10.1023/A:1022193728205

Taylor, K.M., & Rohrer, D. (2010). The effects of interleaved practice. *Applied Cognitive Psychology, 24*, 837-848. http://doi.org/10.1002/ACP.159

Trowler, V. (2010). Student Engagement Literature Review. The Higher Education Academy, 11. 1-15.

Tselios, N., Avouris, N. & Komis, V. (2008). The effective combination of hybrid usability methods in evaluating educational applications of ICT: Issues and challenges. *Educ Inf Technol 13*, 55–76. https://doi.org/10.1007/s10639-007-9045-5

Yan, V., Bjork, E., & Bjork, R. (2016). On the difficulty of mending metacognitive illusions: a priori theories, fluency effects, and misattributions of the interleaving benefit. *Journal of Experimental Psychology: General, 145*(7), 918–933. http:/doi:10.1037/xge0000177

Yan, V., & Sana, F. (2020). Does the interleaving effect extend to unrelated concepts? Learners' beliefs versus empirical evidence. *Journal of Educational Psychology, 113*(1), 125-137. http:/doi.org/10.1037/edu0000470

Yi, M., & Davis F. (2003). Developing and validating and observational learning model of computer software training and skill acquisition. *Information Systems Research, 14*(2), 127-219. http://doi.org/10.1287/isre.14.2.146.16016

van der Meij, H., & Gellevij, M. (2004). The four components of a procedure. *Ieee Transactions on Professional Communication*, *47*(1). https://doi.org/10.1109/TPC.2004.824292

van der Meij, H., & van der Meij, J. (2013). Eight guidelines for the design of instructional videos for software training. *Technical Communication, 60*(3), 205-228.

van der Meij, H. (2014). Developing and testing a video tutorial for software training. *Technical Communication 61*(2), 110-122.

van der Meij, H., & van der Meij, J. (2016). Demonstration-based training (DBT) in the design of a video tutorial for software training. *Instructional Science, 44*(6), 527-542. https://doi.org/10.1007/s11251-016-9394-9

van der Meij, H., van der Meij, J., Voerman, T.E., & Duipmans, E. (2017). Supporting motivation, task performance and retention in video tutorials for software training. *Educational Technology Research and Development*, 1-18.

van der Meij, H. (2018). Cognitive and motivational effects of practice with videos for software training. *Technical Communication*, *65*(3), 265-279.

van der Meij, H., Rensink, I., & van der Meij, J. (2018). Effects of practice with videos for software training. *Computers in Human Behavior*, *89*, 439-445. https://doi.org/10.1016/j.chb.2017.11.029.

van der Meij, H., & Maseland, J. (2021). Practice schedules in a video-based software training arrangement. *Social Sciences & Humanities Open, 3* (1). https://doi.org/10.1016/j.ssaho.2021.100133

van Nimwegen, C., van Oostendorp, H., Burgos, D., & Koper, R. (2006). Does an interface with less assistance provoke more thoughtful behaviour? In Barab, S. A., Hay, K. E., & Hickey, D. T. (Eds.), The International Conference of the Learning Sciences: Indiana University 2006. Proceedings of ICLS 2006, Volume 2 (pp. 785-791). Bloomington, Indiana, USA: International Society of the Learning Sciences. https://doi.dx.org/10.22318/icls2006.785

Waddell, T., Zhang, B., & Sundar, S. (2015). Human–Computer Interaction. The International Encyclopedia of Interpersonal Communication, page 1-9, http:doi.org/10.1002/97811185440190.wbeic182

Zhang, J., & Norman, D. (1994). Representations in distributed cognitive tasks. *Cognitive Science, 18*(1), 87-122. https://doi.org/10.1016/0364-0213(94)90021-3

Appendix A

Web-links to the Graasp spaces

- Blocked practice condition
  https://graasp.eu/s/5n1h4m
- Interleaved practice condition
  https://graasp.eu/s/8q9c2f
- Cumulative practice condition
  https://graasp.eu/s/nxtdtp

Appendix B

Experimental procedures by practice conditions

Blocked condition	Interleaved condition	Cumulative condition
Introduction into the study	Introduction into the study	Introduction into the study
Self-efficacy questionnaire	Self-efficacy questionnaire	Self-efficacy questionnaire
Procedural knowledge pretest	Procedural knowledge pretest	Procedural knowledge pretest
Knowledge pretest	Knowledge pretest	Knowledge pretest
Video Introduction	Video Introduction	Video Introduction
Video 1. Creating acronyms	*Video 1. Creating acronyms*	*Video 1. Creating acronyms*
Practice: Creating acronyms	*Video 2. Autotext usage*	Practice: Creating acronyms
Practice: Creating acronyms	*Video 3. Creating non-breaking space*	*Video 2. Autotext usage*
Video 2. Autotext usage	Practice: Autotext usage	Practice: Creating acronyms
Practice: Autotext usage	Practice: Creating acronyms	Practice: Autotext usage
Practice: Autotext usage	Practice: Creating non-breaking space	*Video 3. Creating non-breaking space*
Video 3. Creating captions	Practice: Autotext usage	Practice: Autotext usage
Practice: Creating captions	Practice: Creating acronyms	Practice: Creating non-breaking space
Recommended 30-second break	Practice: Creating non-breaking space	Recommended 30-second break
Video 4. Cross-referencing to a figure	Recommended 30-second break	*Video 4. Cross-referencing to a figure*
Practice: Cross-referencing to a figure	*Video 4. Creating captions*	Practice: Creating non-breaking space
Practice: Cross-referencing to a figure	*Video 5. Cross-referencing to a figure*	Practice: Cross-referencing to a figure
Video 5. Creating non-breaking space	*Video 6. Keeping header and paragraph together*	*Video 5. Keeping header and paragraph together*
Practice: Creating non-breaking space	Practice: Cross-referencing to a figure	Practice: Cross-referencing to a figure
Practice: Creating non-breaking space	Practice: Keeping header and paragraph together	Practice: Keeping header and paragraph together
Video 6. Keeping header and paragraph together	Practice: Creating captions	Video 6. Creating captions
Practice: Keeping header and paragraph together	Practice: Keeping header and paragraph together	Practice: Keeping header and paragraph together
Practice: Keeping header and paragraph together	Practice: Cross-referencing to a figure	Practice: Creating captions
Recommended 60-second break	Recommended 60-second break	Recommended 60-second break
Self-efficacy and preference questionnaire	Self-efficacy and preference questionnaire	Self-efficacy and preference questionnaire
Procedural knowledge post-test	Procedural knowledge posttest	Procedural knowledge posttest
Conceptual knowledge posttest	Conceptual knowledge posttest	Conceptual knowledge posttest

Appendix C

An extract from the practice Word document

Task 5. Create a Caption under the figure below to ensure the accuracy of Figures. Call the Caption "Figure 1. Video tutorial for beginners".

The Adobe website for Premiere Pro tutorials can be used to illustrate what videos are on offer (Adobe, 2017, August 1). The Figure 1 shows a section from the website's homepage with access to five videos for beginners. Each video link shows a screenshot, a title and a characterization of user activity (i.e., "Watch" or "Try it"), and video duration. The website presents three different formats as possible ways to provide instructions about software usage in a video tutorial: *Embedded, Stand- alone View, Stand-alone Try-it.*

Task 6. In the following text, after the word *see* in the parenthesis, (highlighted with green), create a Cross-reference to the *Figure* below the text.

Before the video opens, the website draws the user's attention to the possibility of downloading practice files as opportunity for practice (see). The video also mentions this possibility. As far as we have been able to establish, this approach is not unique for Adobe. TechSmith also complements its tutorial videos with links to practice files, and there are likely to be other software makers doing so as well. It is this coupling of video and practice that is focal in this paper.

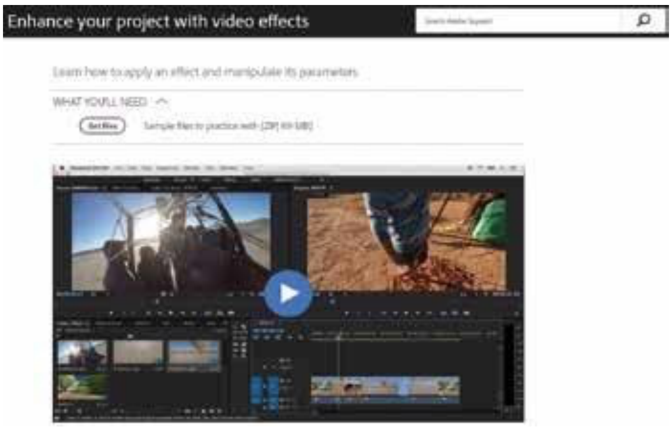
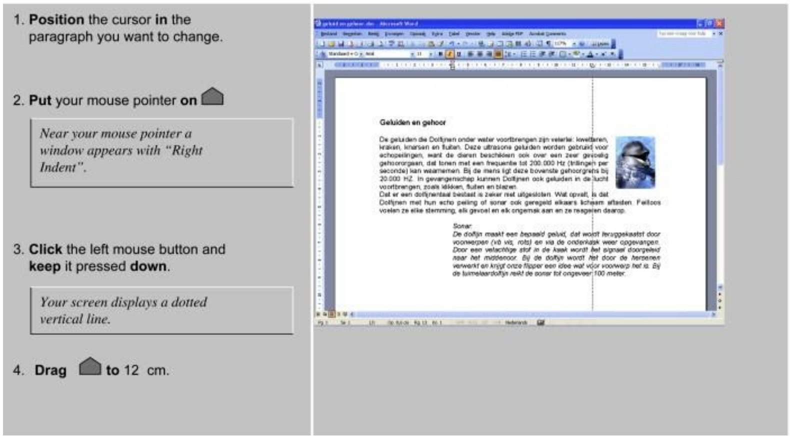


Figure 1. An Adobe's video tutorial example

Appendix D

An extract from the procedural knowledge test

Task 3. Assume that the figure below is the first figure in the document. It has not been numbered yet. Create a Caption under the figure to ensure the accuracy of Figures. Call the Caption “A section from the basic manual”:



Task 4. In the following text, after the word “see” in the parenthesis (highlighted with green), create a Cross-reference to the *Figure* that you see below the text:

In this study, we have tried to further optimize minimalist instructions for user affect with three special measures: relevance organizers, screen displays and deliberate vocabulary use in error-information. Relevance organizers depict a before–after scenario (see).

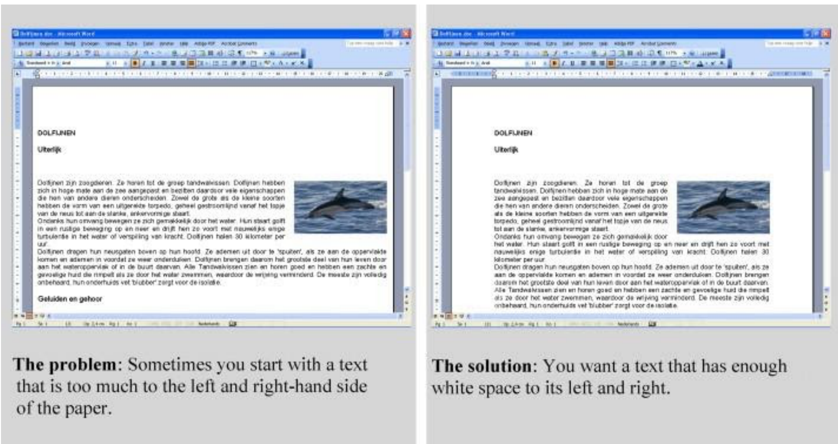


Figure 1. A before-after display to enhance the perception of task-relevance.

Appendix E

The self-efficacy questionnaire

Please answer questions below. Choose the most appropriate answer for you:

Question 1. When a phrase is regularly used in a text, Word allows you to create acronym so that whenever you type the acronym, it is automatically replaced with the full phrase. In your opinion, how well you can perform this task with this option?

Very poor	Poor	Not sure	Well	Very well
☐	☐	☐	☐	☐

Question 2. When you use a phrase frequently in a text, you can have Word suggest the full phrase to you when you type the first few letters of the full phrase by activating the AutoText option.
In your opinion, how well you can perform this task with this option?

Very poor	Poor	Not sure	Well	Very well
☐	☐	☐	☐	☐

Question 3. Word allows you to insert numbers of pages in a Word document (e.g., at the bottom and the right side of each page in a document). In your opinion, how well you can perform this task with this option?

Very poor	Poor	Not sure	Well	Very well
☐	☐	☐	☐	☐

Question 4. Word allows you to automatically assign the correct numbers to all Figures in a text by activating the Create Caption option. In your opinion, how well you can perform this task with this option?

Very poor	Poor	Not sure	Well	Very well
☐	☐	☐	☐	☐

Question 5. Word allows you to refer to a certain Figure in the text of a document by activating the Cross-reference option. In your opinion, how well you can perform this task with this option?

Very poor	Poor	Not sure	Well	Very well
☐	☐	☐	☐	☐

Question 6. When a statistical formula breaks off in the end of a line in a text, Word allows you to keep all elements of the formula together by creating Nonbreaking Spaces. In your opinion, how well you can perform this task with this option?

Very poor	Poor	Not sure	Well	Very well
☐	☐	☐	☐	☐

Question 7. There is the Margins option in Word that allows you to adjust margins of the whole text to the required margin sizes (e.g., to Normal margins). In your opinion, how well you can perform this task with this option?

Very poor	Poor	Not sure	Well	Very well
☐	☐	☐	☐	☐

Question 8. Sometimes you find a header of a paragraph at the bottom of one page whereas the paragraph itself is found on the next page. Word allows you to activate an option in a menu to keep headers and corresponding paragraphs together throughout a text. In your opinion, how well you can perform this task with this option?

Very poor	Poor	Not sure	Well	Very well
☐	☐	☐	☐	☐

Question 9. Word gives you an option to style headings in a text, in accordance with their levels. For instance, a heading can be styled as Heading 1. In your opinion, how well you can perform this task with this option?

Very poor	Poor	Not sure	Well	Very well
☐	☐	☐	☐	☐

Appendix F

Compatibility of MS Word processing tasks in mixed practice schedules

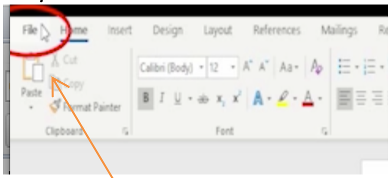
MS Word tasks	Creating an acronym	Autotext usage	Creating captions	Cross-referencing to a figure	Creating non-breaking space	Keeping header and paragraph together
Creating an acronym	x	v	v	v	x	x
Autotext usage	v	x	v	x	v	v
Creating captions	v	v	x	x	v	v
Cross-referencing to a figure	v	x	x	x	v	v
Creating non-breaking space	v	v	v	v	x	v
Keeping header and paragraph together	v	v	v	v	v	x

Note: x - not recommended for delivering in a close proximity to each other in a mixed practice set, v - are unlikely to impede CI-effect if delivered in a close proximity to each other in a mixed practice set

Appendix G

The command map for the task "Using acronyms to shorten typing core phrases"

Step 1

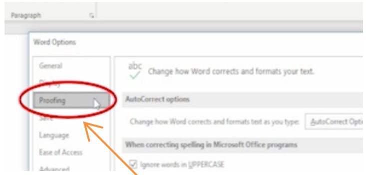


Click File

Step 2

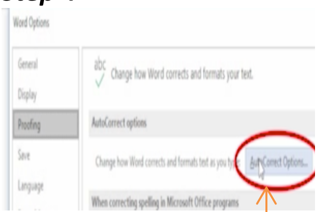
Click 'Options'

Step 3



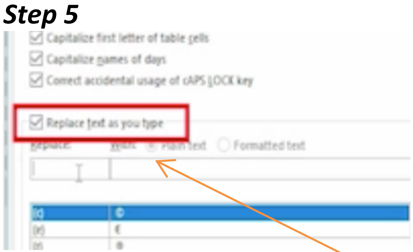
A new window opens up.
Click 'Proofing'

Step 4

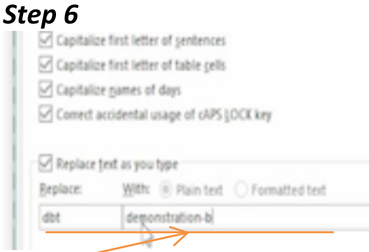


click 'AutoCorrect Options'

Step 5

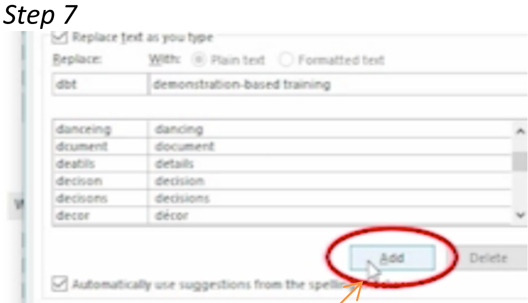


Step 6



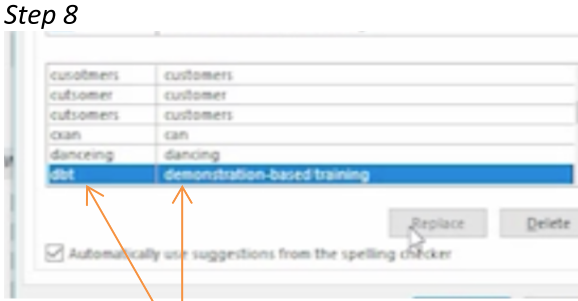
a new window opens up. Below the choice 'Replace text as you type fill in the options for Replace and With. In Replace you type the acronym "dbt". In the empty With cell, you type "demonstration-based training".

Step 7



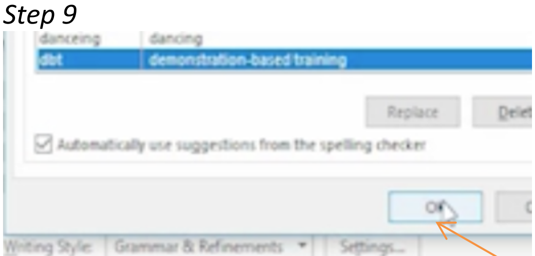
Click 'Add'

Step 8

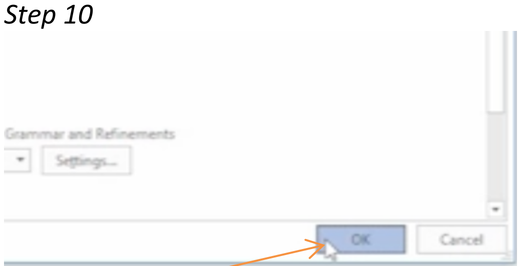


You can see that " dbt " appears under Replace and that it will be substituted for "demonstration-based training"

Step 9



Step 10



Click 'OK' *and another* 'OK' *to go back to your document.*

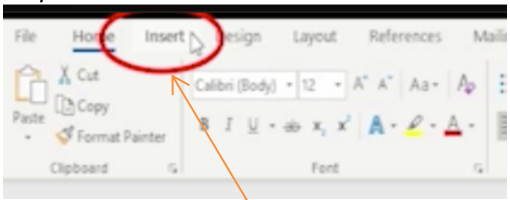
Appendix H

The command map for the task "Using AutoText to have Word suggest phrases"

Step 1

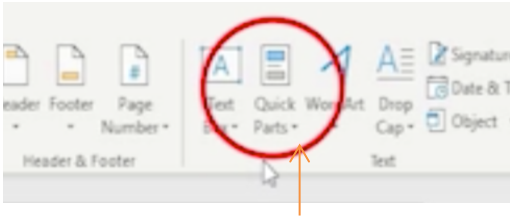
select the full phrase

Step 2



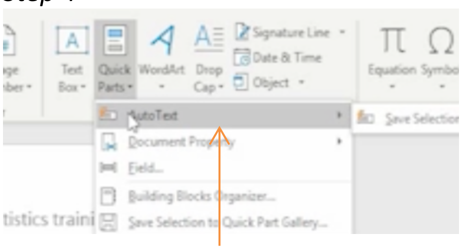
click 'Insert'

Step 3



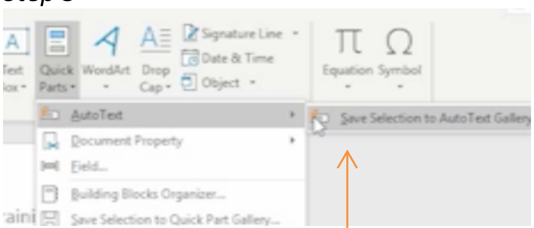
click 'Quick Parts'

Step 4



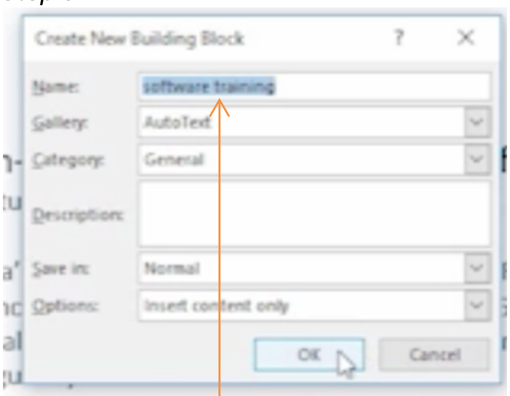
A dropdown menu appears.
Click AutoText

Step 5

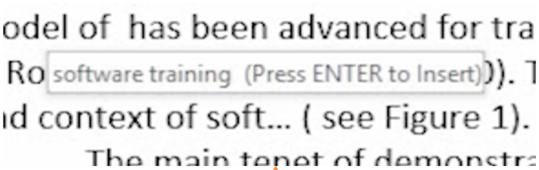


click 'Save selection to AutoText Gallery'

Step 6



In this pop-up window you can see that the phrase you selected is now an Autotext. Click OK *to save it.*

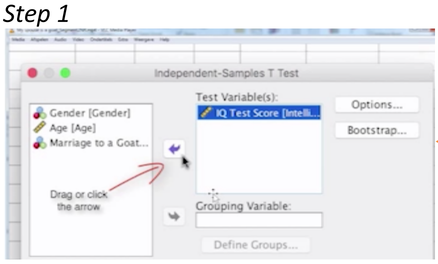


Over here, we want to type the phrase software training again. So we type soft and Press Enter

Appendix I

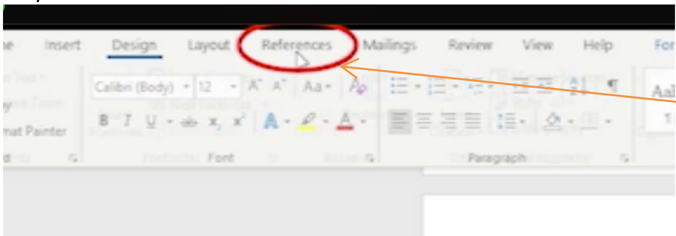
The command map for the task "Creating captions in a MS Word document: numbering and labelling figures"

Step 1



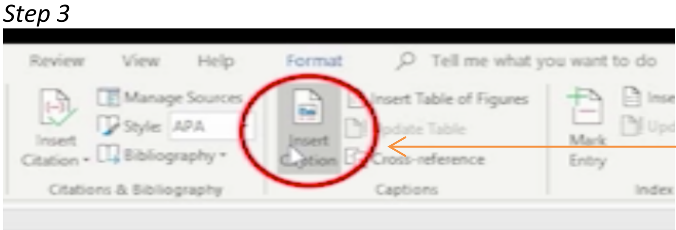
"*First, select* Figure 2…"

Step 2



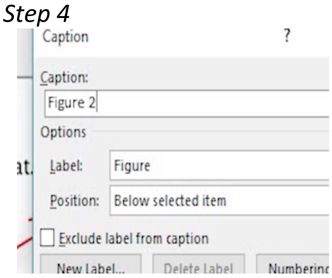
"*Click* REFERENCES"

Step 3

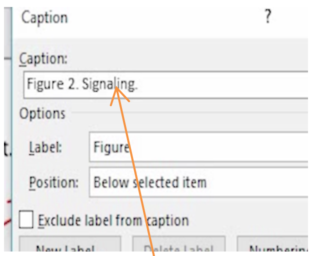


"*Click* INSERT CAPTION*"*

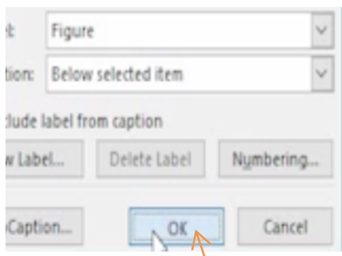
Step 4



A new window opens up

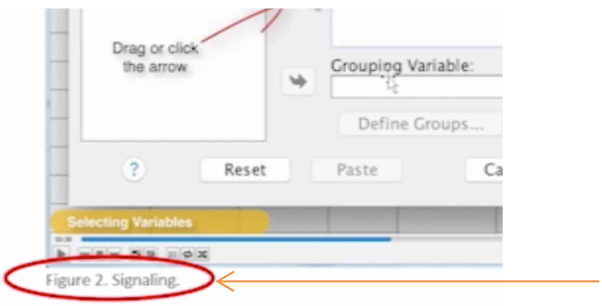


"…type Signaling and end the sentence with a dot."



"Click OK"

Step 5

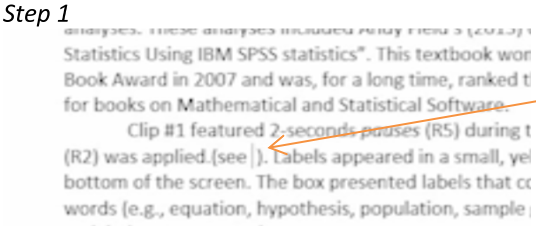


You can now see that the figure caption has appeared

Appendix J

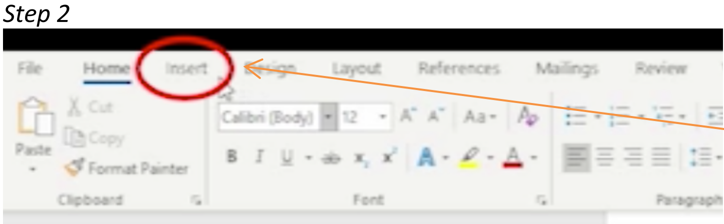
The command map for the task "Cross-referencing in a MS Word document: creating a hyperlink to refer to a figure"

Step 1



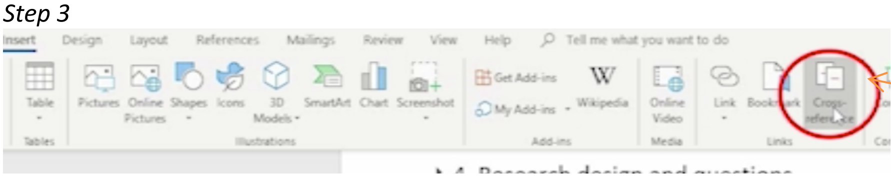
«Move to where figure should be mentioned in the text»

Step 2



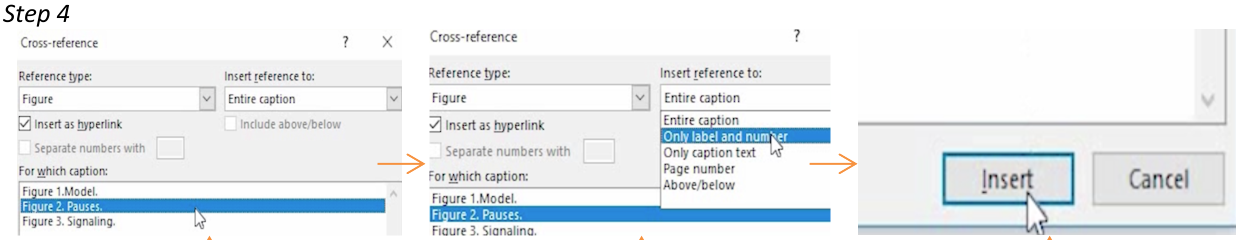
«Click INSERT»

Step 3



«Click Cross-Reference»

Step 4

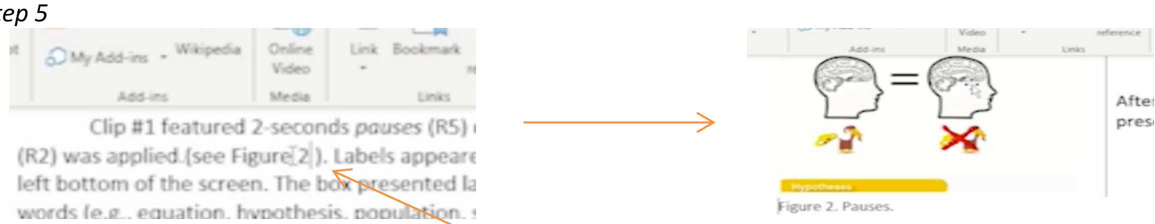


«A new window appears... Indicate that you refer to Figure 2»

«This option should state only Label and Number»

«Click INSERT»

Step 5

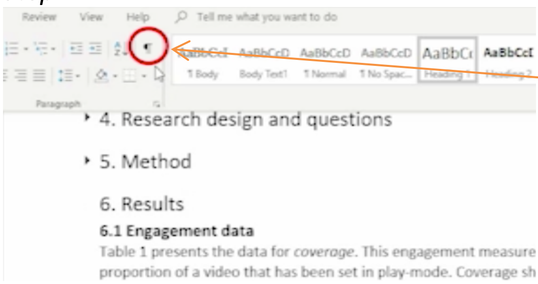


«Press CONTROL *to see how cross-reference jumps to Figure 2*

Appendix K

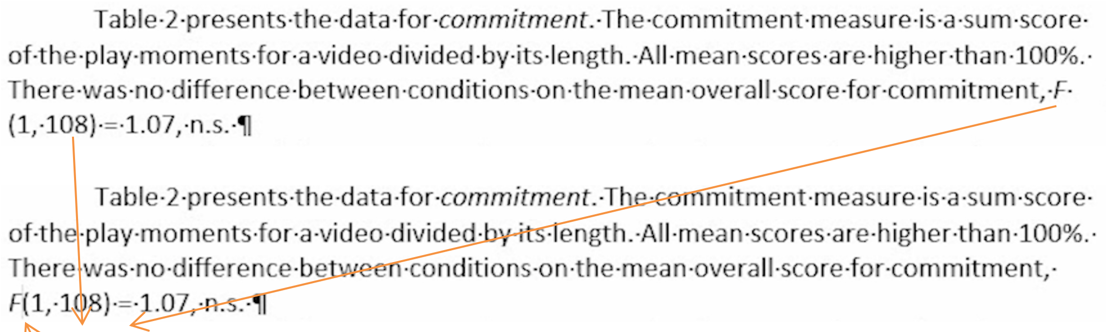
The command map for the task "Nonbreaking space: eliminating spaces in formulas split between two lines in a text"

Step 1



"*click* the Show and Hide *button*"

Step 2

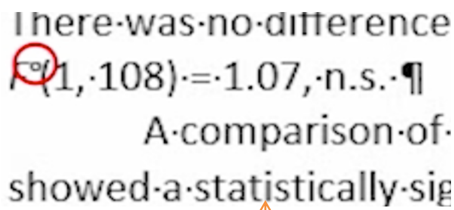


"To create a nonbreaking space, first remove the ordinary space"

Step 3

"Next simultaneously press the keyboard combination of [Ctrl]+[Shift]+[Space]*"*

Step 4



"This is now a nonbreaking space. You must repeat this process until all parts of your formula are connected."

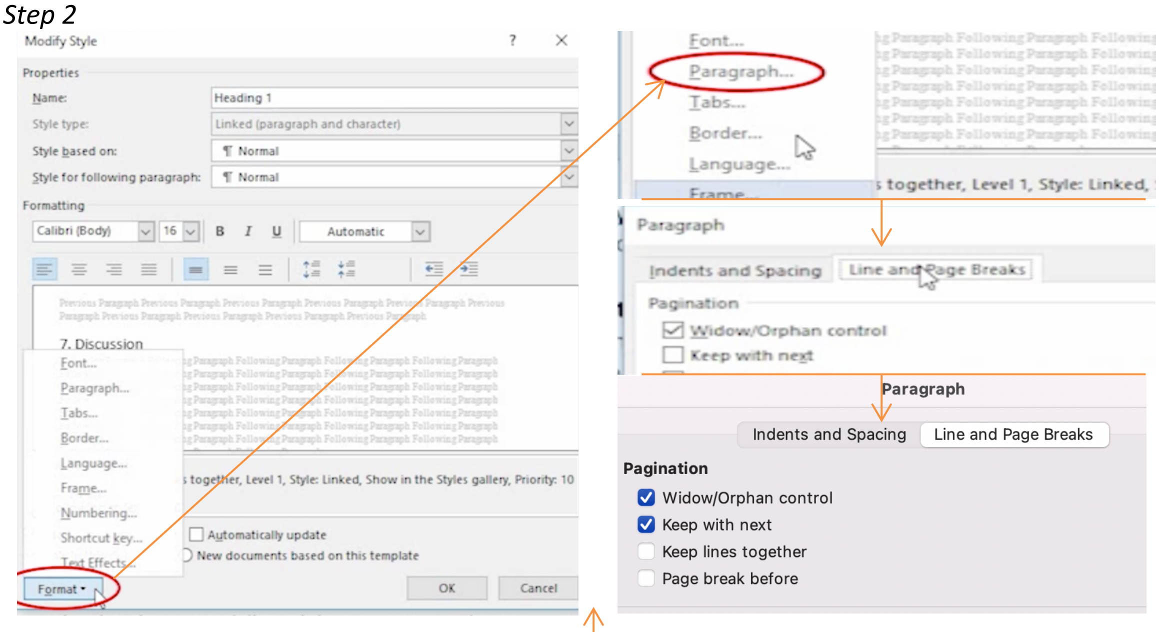
Appendix L

The command map for the task "Keeping together header and paragraph"

Step 1

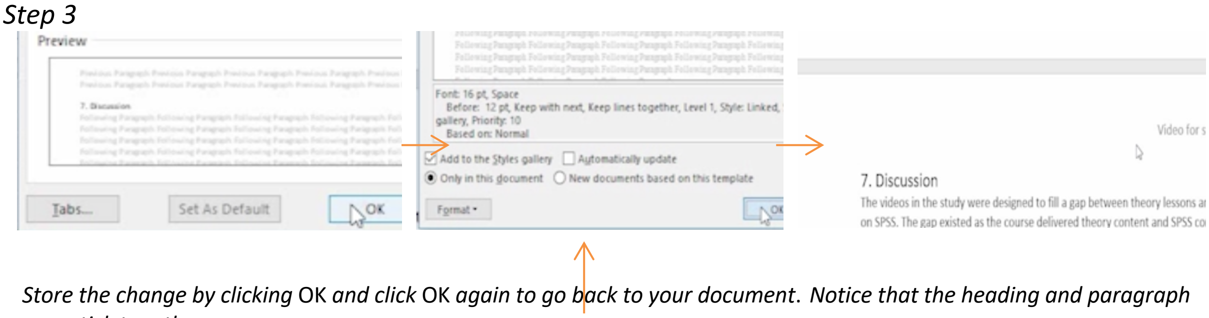
"First, right click on Heading 1."   *Then, click* Modify.

Step 2



"A new window pops up. Click FORMAT *and then click* PARAGRAPH. *A window with two tabs appears*. Click 'Line and Page breaks'. *Here, click* 'Keep with next'."

Step 3

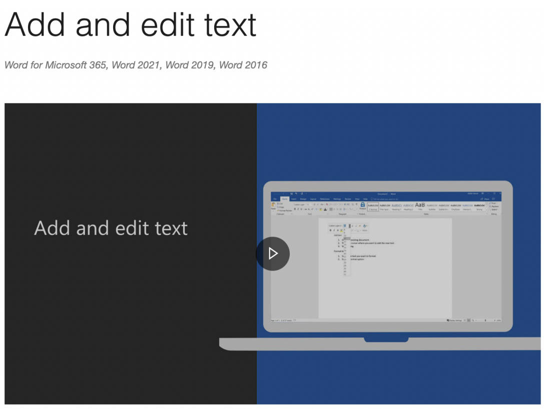


Store the change by clicking OK *and click* OK *again to go back to your document. Notice that the heading and paragraph now stick together.*

Appendix M

An example of existing video-based software training provided on software developers' websites

An instructional component of training:



A practice component of training:

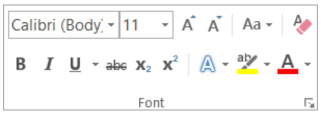
Try it!

Follow these steps to add, replace, and format text in Word.

Add text
Place the cursor where you want to add the text.
Start typing.

Replace text
Select the text you want to replace.
To select a single word, double-click it.
To select a line, click to the left of it.
Start typing.

Format text
Select the text you want to format.
Select an option to change the font, font size, font colour, or make the text bold, italics, or underline.



Copy formatting
Select the text with the formatting you want to copy.
Click **Format painter**, and then select the text you want to copy the formatting to.

Source: https://support.microsoft.com/en-us/office/add-and-edit-text-ed1e3147-a846-41ca-8087-49e324cb50bd?wt.mc_id=otc_word#