



Medical education videos as a tool for rehearsal: efficiency and the cases of background music and difficulty

Hauke S. Meyerhoff^{1,2} · Martin Merkt³ · Carla Schröpel^{2,4} · Adrian Meder⁵

Received: 15 June 2021 / Accepted: 3 July 2022 / Published online: 2 September 2022
© The Author(s), under exclusive licence to Springer Nature B.V. 2022

Abstract

This study reports a field experiment investigating how instructional videos with and without background music contribute to the learning of examination techniques within a formal curriculum of medical teaching. Following a classroom teaching unit on the techniques for examining the knee and the shoulder joint, our participants ($N=175$) rehearsed the studied techniques for either the knee or the shoulder joint with an instructional video with or without background music. As dependent measures, we collected a general questionnaire, a prediction of test performance, as well as performance on an exam-like knowledge test covering both joints. For both videos, the participants who had watched the particular video during rehearsal were more accurate in answering the corresponding questions than the participants who had seen the other video, signaling that instructional videos provide a useful tool for rehearsal (i.e., both groups reciprocally served as control groups). For the knee video (less difficult), we observed a detrimental effect of the background music, whereas we observed no such effect for the shoulder video (more difficult). Further explorations revealed that background music might be detrimental for learning, as it reduces the perceived demand characteristics. Because the impact of the demand characteristics might be more pronounced in less difficult instructional videos, we discuss video difficulty as a potential moderating factor. Overall, our study provides evidence that instructional videos could be usefully implemented in formal teaching curricula and that such instructional videos probably should be designed without background music.

Keywords Learning · Rehearsal · Instructional video

✉ Hauke S. Meyerhoff
hauke.meyerhoff@uni-erfurt.de

¹ University of Erfurt, Nordhäuser Straße. 63, 99089 Erfurt, Germany

² Leibniz-Institut für Wissensmedien, Tübingen, Germany

³ German Institute for Adult Education, Bonn, Germany

⁴ Psychosomatic Medicine and Psychotherapy, University Hospital Tübingen, Tübingen, Germany

⁵ Berufsgenossenschaftliche Unfallklinik, University of Tübingen, Tübingen, Germany

Introduction

Using instructional videos in medical education is appealing because – once edited – they seemingly offer an efficient way of communicating knowledge especially because they can be disseminated to a large number of learners in a spatially and temporally flexible manner (Kay, 2012; Kay & Kletschin, 2012). However, the success of instructional videos in terms of knowledge acquisition cannot be taken for granted (Salomon, 1984; Kardas & O'Brien, 2018). Instead, successful teaching with instructional videos requires careful consideration about the design of the video (e.g., Biard et al., 2018; Fiorella & Mayer, 2018; Hoogerheide et al., 2016; Merkt et al., 2020). In the present report, we follow a use-inspired basic research approach which aims at investigating basic research questions in the applied context of medical teaching. First, we examine whether instructional videos can be considered an effective tool for rehearsal of previous curricular teaching. Second, we investigate if background music alters knowledge acquisition from the rehearsal videos in this applied context, as previous research has critically and controversially discussed beneficial versus detrimental effects of background music in educational psychology and related fields of research (e.g., Boeckmann et al., 1990; Kopiez et al., 2013; Lehmann et al., 2019).

Opportunities and Challenges of Instructional Videos

In the current research project, we embedded an instructional video within a formal curriculum of medical training. Within this curriculum, watching the video served as rehearsal, reflecting an important aspect of long-term learning, in particular with variability between the learning materials across repetitions (e.g., Bjork & Bjork 1992). The two most evident differences between instructional videos and more traditional learning formats such as textbooks are the multimodal nature of instructional videos as well as the dynamic presentation of the materials to be learnt. On the one hand, the multimodal presentation of videos or animated stimuli has been demonstrated to be beneficial for memory (Meyerhoff & Huff, 2016) and learning (i.e., the modality principle; Mayer 2001; or the modality effect; Sweller et al., 1998), mostly because learners could use their (limited) working memory resources more efficiently when the relevant information was distributed across modalities. In contrast, the dynamic presentation of information in instructional videos might provide an obstacle for learning, because missed information cannot be revisited without further costs of interacting with the video (Ayres & Paas, 2007; Schwan & Riempp, 2004). While this transience might limit the usefulness of instructional videos for some domains of knowledge, the dynamic nature of the video-based presentation might be particularly useful for other domains. Borrowing evidence from research on instructional animations, a recent systematic review has identified procedural knowledge to particularly benefit from animated instructional materials (Ploetzner, Berney, & Bétrancourt, 2020, also see Höffler & Leutner 2007). Given the structural similarities between animations and instructional videos, we argue that procedural knowledge is also likely to benefit from the dynamic nature of instructional videos.

Perceived Demand Characteristics

Beyond considerations about design characteristics of instructional videos, Salomon (1984) has identified a general challenge for knowledge acquisition with videos, which emerges from perceived demand characteristics. In his study, he asked pupils to either learn with a silent video or a matching printed text. The challenging aspect of the video format was that it was associated with lower *perceived demand characteristics*, which resulted in superficial processing of the learning materials. Therefore, the participants in the text condition acquired more knowledge than those in the video condition. The low perceived demand characteristics of videos may thus provide an inherent problem of the presentation format for educational purposes. Further research has confirmed Salomon's results by observing memory advantages of printed materials relative to television-like presentation modes for scientific information (e.g., Furnham et al., 1990) or news contents (e.g., DeFleur et al., 1992; Gunter et al., 1984). Further, in line with Salomon's argument, recent experiments by Kardas & O'Brien (2018) demonstrated that repeated exposure to video-based learning materials actually increased the learners' confidence in being able to perform the corresponding task without actually boosting their mastery.

The concept of perceived demand characteristics is related to more recent work that has studied the impact of perceived fluency on perceived learning as well as actual learning. For instance, the fluency of an instructor within a video (i.e., a well vs. a poorly prepared instructor) increased the subjective impression of learning which was reflected in higher predictions of test performance. Critically, however, this impression was not accompanied by an actual increase in learning (Carpenter et al., 2013; Toftness et al., 2018), showing that the participants had overrated their own learning success (see also Kornell & Bjork 2008). With regard to actual learning, however, related research has argued that experimental conditions introducing disfluency could be beneficial for the learning outcome (e.g., Diemand-Yauman et al., 2011; but see Kühl et al., 2014). Such an effect might well explain detrimental effects of videos on learning relative to textual presentations. In particular, the rich visual and auditory possibilities of videos could create an impression of perceptual fluency, which tends to be misattributed to successful learning, thus creating an illusion of knowledge (Ryffel & Wirth, 2020). Therefore, if videos are to be used because they are more suitable for imparting a certain type of knowledge (i.e., procedures; see Ploetzner et al., 2020), it is important to avoid conditions that may additionally lower the perceived demand characteristics of videos. One candidate characteristic that might lower the perceived demand characteristics of an instructional video is background music which could be implemented within the video.

Background Music

The impact of background music on learning has been controversially discussed within the literature (see de la Mora Velasco & Hirumi 2020, for a review), ranging from general improvements in cognitive ability (Rauscher et al., 1993), to an increase in learning relevant cognitive activities (Lehmann et al., 2019), to considering its influence to be mostly negligible (Kopiez et al., 2013) or even harmful for learning (e.g., Boeckmann et al., 1990; Brosius, 1990; Lehmann & Seufert, 2017). Given the wide variation in tasks, involved mental

processes, as well as diverse choices of background music, such conflicting result patterns appear to be unavoidable.

In addition, different theories lead to different predictions about the effects of background music in learning materials. On the one hand, from the viewpoint of the cognitive theory of multimedia learning (Mayer, 2001) or the cognitive load theory (Sweller et al., 1998), the value of background music is questionable at best. This is because background music typically does not transport any relevant information for learning but, instead, might distract from relevant information or learning activities (Mayer, 2001; Sweller et al., 1998). This reasoning is in line with an expert analysis of medical education videos on YouTube reported by Azer et al., (2022). On the other hand, numerous research approaches have provided alternative theoretical explanations of how background music could potentially have beneficial effects on learning. These explanations focus on mood (Boltz et al., 1991; Husain et al., 2002), arousal (Husain et al., 2002; Thompson et al., 2011; Lehmann et al., 2019), or valence (Eschrich, Münter, & Altenmüller, 2008).

Direct attempts to evaluate the impact of background music on learning are rare and have revealed inconsistent results. Boeckmann et al., (1990) varied the background music (music designed for learning vs. supermarket music vs. attentionally alerting music) in three educational film clips and contrasted these film clips to versions without background music. With regard to the recall of presented information, the results varied across the different film clips and the presentation modality of the tested information. On average, however, background music impaired rather than increased recall performance, and the type of background music only had negligible effects. A similar study was conducted by Brosius (1990) who also investigated the effect of information presented in informational film clips. In this study, the background music either matched the original soundtrack, opposed this soundtrack in terms of its features such as speed, or was absent. The results of this study were also mixed. Nevertheless, for one of the films, background music impaired recognition performance but boosted appraisal of the video. Motivated by the discrepancies observed in the preceding studies, Kopiez et al., (2013) conducted a systematic investigation of the effects of background music for informative film clips from news magazines. The background music was orthogonally manipulated along the axes of arousal and valence. In contrast to the previous studies, Kopiez et al., (2013) did not observe a substantial impact of the background music on recognition performance for the presented information.

Despite the lack of direct evidence for a beneficial effect of background music in instructional videos on learning, instructional videos with background music are more popular than those without background music (ten Hove & van der Meij, 2015). It thus seems likely that this increase in popularity arises from processes which are not directly linked to learning or might even be harmful for actual learning. One possibility is that background music increases the perceptual fluency of the video. Direct evidence for this suggestion emerges from studies demonstrating a more pronounced edit blindness (i.e., missing filmic cuts even when explicitly asked to identify them) when an auditory track is added to a visual track (Meitz et al., 2020; Smith & Martin-Portugues Santacreu, 2017). As related impressions of fluency have been demonstrated to elicit illusions of learning (Carpenter et al., 2013; Ryyfel & Wirth, 2020), a reduction in invested mental effort following illusory impressions of learning appears to be a candidate mechanism for a reduction of the learning outcome in some of the previous studies (Boeckmann et al., 1990; Brosius, 1990).

Current Study and Research Questions

Based on the considerations about videos as a potential tool for rehearsal as well as the unclear impact of background music, we decided to combine these two fields of research in this project. A central caveat from previous work is that the inconsistent result patterns make it difficult to generalize design principles for instructional videos in applied settings such as actual teaching. Addressing this challenge, we tested the efficiency of instructional videos with and without background music in an applied setting. Because dynamic procedures such as examination techniques represent a central pillar in the field of medical education (Ruessler et al., 2010), and previous research has identified such dynamic processes to be a good candidate for teaching with dynamic representations (Ploetzner et al., 2020), we ran a field experiment within the medical curriculum. This approach builds on previous evidence that learning with videos is capable of increasing procedural knowledge in the medical disciplines (Boucheix et al., 2018).

For our study, we therefore implemented a brief rehearsal session (approx. 2 h after initial teaching) within the medical teaching curriculum. Within this rehearsal session, we presented the participants one of two videos (knee vs. shoulder joint). Orthogonal to the content of the video, one half of the participants saw the rehearsal video with background music, whereas the other half saw the rehearsal video without background music. Importantly, the background music was selected by a professional filmmaker of the University of Tübingen in order to cover the breaks between passages with and without verbal instruction but did not include a melody to avoid distraction from or interference with the study materials. Because it was impossible to match the instructional videos for difficulty in our applied setting, we decided to implement two videos which differed substantially enough in their difficulty so that we could explore the effect of difficulty on the success of rehearsal. This was motivated by the possibility that the inconsistent result patterns in previous research may have emerged from differences in task difficulty. Our main dependent variable was the score in an exam-like knowledge test which covered the topics of both instructional videos (i.e., the seen as well as the unseen video; the questions of the unseen video served as control condition). Additionally, the participants provided a prediction score regarding their expected performance prior to the knowledge test and completed a questionnaire with which we aimed to gain further insight into the processes driving our results. This questionnaire covered the evaluation and appraisal of the instructional video as well as measures of cognitive load, perceived demand characteristics, and invested mental effort.

With regard to the test scores, we expected to observe a benefit resulting from rehearsal. We refer to this as the *rehearsal hypothesis*. Further, we intended to investigate the effect of the background music on test scores. We refer to this as the *background music hypothesis*. Given the conflicting evidence within the literature as well as our attempts to use background music in the least obtrusive manner, a clear a priori prediction for this hypothesis was difficult. Whereas a positive impact of the background music seemed unlikely, both a negative impact (as in Boeckmann et al., 1990; Brosius, 1990), or a negligible impact (as in Kopiez et al., 2013) were in the range of realistic outcomes. Finally, the impact of video difficulty served mostly exploratory purposes without clear theoretical predictions.

With regard to the participants' prediction of their scores in the knowledge test, we were interested in whether background music results in an illusion of learning, meaning that the participants tend to estimate conditions as beneficial for learning which are actually irrel-

evant or even harmful. We refer to this as the *learning prediction hypothesis*. We tested this hypothesis by comparing the result patterns of predicted learning with the result pattern of the actual test scores, depending on whether the participants had watched the video with or without background music.

With the questionnaire, we were interested in providing further evidence on how background music affects test performance and the mechanisms behind this effect. Comparable to the difficulty manipulation, this was implemented mostly for exploratory reasons and to outline informed paths for future research.

Methods

Participants

The sample consisted of 175 students (101 female, 73 male, 1 unanswered). According to a power analysis, a sample size of 152 participants would have been sufficient to observe the main effects of interest (between-subject) in our experiment with an effect size of $\eta^2_p > 0.05$ and a power of $(1 - \beta) > 0.80$ at $\alpha = 0.05$. Nevertheless, we decided to test the entire semester in order to compensate for potentially weaker manifestations of the effects. All students were enrolled in the 9th semester of the medical curriculum at the University of Tübingen, Germany. The curriculum is highly structured, so that there is hardly any variation between the students in terms of attended classes. The study was conducted during an obligatory¹ one-day internship in surgery² which was held in a semester in which most internships were canceled or postponed due to the COVID-19 pandemic (the only other exception was an internship in gynecology which thematically is completely unrelated to the content of our study materials). In the curriculum, the only potential sources of prior knowledge with regard to the knee and shoulder joints were two obligatory lessons (90 min each) in the 2nd semester (i.e., 3.5 years before the study). Most importantly, however, the videos in our study served as a rehearsal of the examination techniques which had been taught within the internship two hours before the study which likely equalizes all participants in terms of prior knowledge.

The age of the participants ranged from 22 to 39 years (three participants did not indicate their age) with a mean age of $M = 26.12$ years ($SD = 3.09$ years). The participants completed the experiment individually but in a group setting with group size varying between 8 and 16 participants. The experiment was conducted in a lecture hall, and participants were seated apart so that they were unable to see each other's responses (and to adhere to the social distancing rules during the COVID-19 pandemic). The experiment took course over a period of 12 weeks with one group of participants each week. The full groups of participants each week were randomly assigned to the four experimental conditions with the overall restriction that three groups had to be assigned to each condition. In return for their participation, participants could sign up for a raffle of five 25€ vouchers. The experimental procedure was approved by the institutional review board of the Leibniz-Institut für Wissensmedien,

¹ The participants had the opportunity to opt out of the study. In this case, they still were able to stay in the room and watch the rehearsal video.

² Beyond the examination techniques for knee and shoulder joint, this internship covered the topics osteosynthesis, surgical suture, and polytrauma.

Tübingen (IWM), Germany (LEK-2020/026), and all participants signed informed consent prior to the study.

Apparatus

The videos were presented using the standard presentation system of the lecture hall, including the loudspeakers for the auditory track. The video was clearly visible, and the auditory track was clearly audible from all seats within the hall. The participants' answers were collected using a paper-pencil questionnaire.

Stimuli & Materials

The stimuli and materials of this study consisted of two educational videos as well as a questionnaire and a knowledge test. Each participant saw one of the videos. The answers to the knowledge questions of the other video served as control condition.

Videos

We used two distinct videos presenting examination techniques for joint injuries in a realistic clinical scenario. These videos emerged from a collaboration between medical educators, educational psychologists, and professional filmmakers. The videos were intended to introduce evidence-based examination techniques for joint injuries (see Fig. 1 for exemplary screenshots).

The first video (14:02 min) addressed examination techniques for the knee joint. Following an introduction of the anatomy of the knee (2:33 min), the video sequentially introduced 20 examination techniques (divided into eight chapters) for injuries of the knee joint. Each chapter, as well as each examination technique, started with a brief introductory screen. For the examination techniques, this introductory screen also displayed the empirical evidence (i.e., sensitivity and specificity) as well as a schematic illustration highlighting the relevant part of the joint for the following technique. In all conditions, the videos were presented at a normal speed that matched the speed from a real examination situation. The techniques were demonstrated by a male examiner with a real female patient. A male off-screen narrator (speaking German) explained all demonstrated examination techniques, including the corresponding criteria for diagnoses. Throughout the demonstration of the examination techniques, the chapter, the technique, as well as the supporting empirical evidence were displayed in the lower left corner of the screen. This video is published with a full table of contents in Meder et al., (2021a).

The second video addressed examination techniques for the shoulder joint. This video was developed following the same basic scheme as the first video but substantially differed in length (19:11 min; 1:40 min introduction to the anatomy of the shoulder) as well as in the number of illustrated examination techniques (36 examination techniques; divided into seven chapters). The techniques were demonstrated with a real male patient. The examiner as well as the voice explaining the examination techniques and all other features of the video were identical to the first video. This video is published with a full table of contents in Meder et al., (2021b).

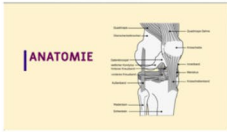


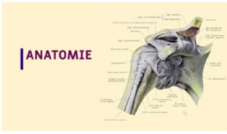


Video Content	General Introduction of the relevant Joint	Introduction of the Examination Technique	Illustration of the Examination Technique
Knee Joint			
Shoulder Joint			

Fig. 1 Exemplary screenshots from the two videos illustrating the examination techniques for the knee joint (upper row) and the shoulder joint (lower row). The second column represents the general introduction to the anatomy of the corresponding joint at the beginning of the video. The third column represents the introduction of specific examination techniques. The fourth column represents the demonstration of the examination technique with real patients

There were two variants of each video: with and without background music. The background music (which was selected by the professional filmmakers) was a rather slow (64 bpm), purely instrumental, relaxing, and spherical soundtrack (Mel O'Dee-Heiko Klüh, without date; available at <https://www.soundtaxi.com/de/Ambient-Atmosphere/Flight-over-the-region-loop::1947>).

Questionnaire

Immediately following the video, the participants completed a questionnaire with 20 questions. For each question, there were seven response options, ranging from complete disagreement to complete agreement. First, we asked them how much they enjoyed learning with the video (Question 1). Second, we asked the participants to rate the perceived professionalism of the video clip, the auditory track, as well as the visual track (Questions 2–4). Third, we asked the participants for their aesthetic experience while watching the video. These questions asked how much they liked the video overall, how much they liked the auditory track, and how disturbing they experienced the breaks in the spoken auditory track (Questions 5–7). Fourth, we assessed invested mental effort (Question 8), using the item reported in Paas (1992), as well as the experienced demand characteristics (Question 9), using the item from the AIME framework (Salomon, 1984). Both items were adapted to the videos as stimuli. Fifth, we assessed intrinsic, extraneous, and germane cognitive load using the eight items reported in Klepsch et al., (2017), which we also adapted to the videos as stimuli (Questions 10–17). We chose the items of Klepsch et al., (2017) because they provide a faceted and validated way to measure three different types of cognitive load and because they were developed and tested in German language, which also was the current study language. Sixth, we asked participants to self-assess their learning experience with the video. In particular, the participants indicated whether they liked the learning unit with the video, whether they would like to do another video unit, and whether the video motivated them to learn. The full questionnaire (in German) available at <https://osf.io/9u5wr/>.

Knowledge Test and Prediction

In the final part of the study, the participants answered five multiple choice questions addressing the content of each video. The questions were designed by medical teachers and matched those of real exams within the curriculum³. There were five response options for each question. The instructions of the knowledge test informed participants that only one of the response options was correct and participants received one point per question if they only ticked the correct option. This resulted in a maximum score of 5 points for each video. The questions addressed the relationships between the illustrated examination techniques and symptoms of an injury. One half of the participants started with the questions addressing the knee joint, whereas the other half started with the questions addressing the shoulder joint. Before seeing the questions of the knowledge test, the participants were asked to provide an estimate of their test score (i.e., 0–5) for the questions on each joint. All questions of the knowledge test, including the performance prediction (in German), are available at <https://osf.io/9u5wr/>

Procedure

The experiment was embedded as a rehearsal unit within the context of a one-day internship in surgery. Among other topics, this internship covered the topic “injuries of the joints”. Within this section of the internship, the students received an introduction into examination techniques of the shoulder and the knee joint. These examination techniques were first presented by a teacher and then practiced in groups of two students each. Following a further introduction as well as a break (together approximately 2 h), the participants were offered to watch a rehearsal video demonstrating the examination techniques. Depending on the experimental condition, the examination techniques either addressed the knee or the shoulder joint and were presented with or without accompanying background music. As part of the study, they then completed the questionnaire and the knowledge test (for the examination techniques of both joints). We counterbalanced the order of questions in the knowledge test (i.e., knee questions first vs. shoulder questions first) to avoid any confounds between the watched video and the order of the questions. Following the test and a debriefing, the participants were allowed to watch the video demonstrating the examination techniques of the joint that they had not seen as part of the study.

Design and Analysis Plan

For the knowledge test score and the predicted test score, the experiment followed a 2 (video content: knee vs. shoulder joint; between participant) x 2 (background music: present

³Because the knowledge tests regarding the knee video and the shoulder video were highly distinct and covered various diagnostic methods, there is no theoretical reason to calculate Cronbach’s alpha as an indicator for the internal consistency of the knowledge tests. In other words, our knowledge tests were designed in analogy to real study exams in the domain of medicine rather than to assess a general construct that would predict response behavior across all questions. Following this consideration, it is not surprising that Cronbach’s alphas regarding the knowledge tests in this experiment are very low. Therefore, the scores in the knowledge tests in our study reflect a purely quantitative assessment of the number of questions in the relevant domain which the participants were able to answer correctly without inferring to an underlying knowledge construct (please see Stadler et al., 2021, for a discussion why high Cronbach’s alpha in knowledge tests appear to contradict this common conceptualizations of knowledge in exam situations).

vs. absent, between participant) \times 2 (congruency between test questions and video: congruent vs. incongruent; within participant) mixed design. For each of these dependent variables, we ran an ANOVA⁴ for mixed designs, followed by post-hoc *t*-tests in order to resolve interactions if they occurred. Regarding the evaluation of the video and the learning process, only the two between participant factors were relevant. In order to analyze the results of the questionnaire, we first aggregated those items that belonged to a scale. Then we ran 2×2 ANOVAs for between subject designs for each dependent measure. Please note that this is a rather exploratory analysis consisting of a total of 15 tests. We will therefore discuss only those with clear theoretical predictions and report the others in a rather exploratory manner.

For the ANOVA analyses, we report generalized eta-squared as measure of the effect size (η_G^2 , Olejnik & Algina 2003). We chose this measure of effect size to facilitate comparability of the effect sizes across different research designs as the mixed design in our study is not a standard design). As effect size for simple comparisons (i.e., *t*-tests), we report Cohen's *d* for comparisons between participants and the matching standardized difference scores d_z .

Results

All data and analysis scripts are available at <https://osf.io/9u5wr/>.

Test Performance

For the analysis of the test performance, we had to exclude data from a total of 10 participants who missed one page of the test and thus responded to only one half of the test items. Two further participants had one (out of 10) missing response which were counted as "incorrect answers". The resulting test scores are depicted in Fig. 2.

We conducted a mixed model ANOVA with the content of the video (knee vs. shoulder; between-subject), the presence of background music in the video (present vs. absent, between-subject), and the congruency between the questions and the video content (congruent vs. incongruent, within-subject) as the independent variables⁵. For participants who attended to the knee video, the knee questions were congruent with the video, whereas the shoulder questions were incongruent. For participants who attended to the shoulder video, the shoulder questions were congruent, whereas the knee questions were incongruent (the incongruent questions served as control condition). As the dependent variable, we used the

⁴ We conducted ANOVAs although it is difficult to confirm some of their prerequisites such as the normality assumption. Although this critique applies to most research reports conducting ANOVAs, ANOVAs are common practice as they allow for investigating interactions between multiple independent variables and have been demonstrated to be robust with regard to violations of the normality assumption (Schmider et al., 2010).

⁵ Because we counterbalanced the order of the questions during the knowledge test, the effect of the order also could be analyzed. Since this is a rather technical factor which is not of theoretical relevance, we did not add this factor into our main analysis. However, we conducted an exploratory ANOVA in which we included the order of the test items as a fourth independent variable (between-subject). There was a main effect of the order, $F(1, 157) = 5.20$, $p = .024$, $\eta_G^2 = 0.02$ indicating that the participants were more accurate across all questions when the shoulder questions preceded the knee questions rather than vice versa. Most importantly, all potential four-, three-, and two-way interactions involving the order of the questions were clearly not significant, all $F_s(1, 157) < 1$, all $p_s > 0.321$, $\eta_G^2 < 0.01$. As the main effect of the order refers to accuracy in the answers to the questions related to both the seen and the unseen video, this effect likely reflects random variation between the participants which we will not discuss any further.

test score (0–5) for the exam-like knowledge questions for the two videos separately, reflecting the congruency manipulation as a within-subject factor.

Most importantly, this analysis revealed a significant three-way interaction between all independent variables, $F(1, 161)=7.24, p=.008, \eta_G^2=0.02$. With regard to the remaining interactions, the two-way interaction between video content and congruency, $F(1, 161)=37.09, p<.001, \eta_G^2=0.09$, as well as the two-way interaction between background music and congruency were significant, $F(1, 161)=4.10, p=.045, \eta_G^2=0.01$. However, there was no two-way interaction between background music and video content, $F(1, 161)<1, p=.775, \eta_G^2<0.01$. With regard to main effects, congruency reached significance, $F(1, 161)=25.35, p<.001, \eta_G^2=0.07$, but neither video content, $F(1, 161)<1, p=.735, \eta_G^2<0.01$, nor background music did, $F(1, 161)<1, p=.899, \eta_G^2<0.01$.

In order to test the two hypotheses addressing test performance, we further resolved the three-way interaction with follow-up analyses. First, we investigated whether attending to a particular video improves answer accuracy for congruent questions relative to incongruent questions. We conducted separate two-way mixed ANOVAs for the knee video and the shoulder video with background music and congruency as the independent variables. For the knee video (i.e. left panel of Fig. 2), this 2×2 -mixed ANOVA confirmed a significant two-way interaction between background music and congruency, $F(1, 89)=11.38, p=.001, \eta_G^2=0.06$. Further, the main effect of congruency was significant, $F(1, 89)=65.25, p<.001, \eta_G^2=0.26$, whereas the main effect of background music was not, $F(1, 89)<1, p=.762, \eta_G^2<0.01$. Follow-up post-hoc *t*-tests for paired samples confirmed that the effect of congruency was present in the condition with background music, $t(45)=3.42, p=.001, d_z = 0.50$, as well as in the condition without background music, $t(44)=7.97, p<.001, d_z = 1.19$, but that the effect was numerically larger in the condition without background music. For the shoulder video (right panel in Fig. 2), the 2×2 -mixed ANOVA revealed no main effects of congruency, $F(1, 72)=1.49, p=.226, \eta_G^2<0.01$, or background music, $F(1, 72)<1, p=.905, \eta_G^2<0.01$, and no interaction between both factors, $F(1, 72)<1, p=.523, \eta_G^2<0.01$.

Second, we investigated whether participants who had attended to one of the videos could answer the questions congruent to the content of that video more accurately than participants who had attended to the other video. To do so, we conducted separate follow-up ANOVAs for the knee questions and for the shoulder questions with video content as well as background music as the independent variables. For the knee questions (i.e. dark grey bars in Fig. 2), this analysis revealed a main effect of video content, $F(1, 161)=7.87, p=.006, \eta_G^2=0.05$, indicating that participants who attended to the knee video were more accurate in responding to the knee questions than participants who attended to the shoulder video. Further, the main effect of background music approached significance with numerically lower test performance with than without background music, $F(1, 161)=3.51, p=.063, \eta_G^2=0.02$, but there was no interaction between both variables, $F(1, 161)=1.56, p=.213, \eta_G^2<0.01$. Of course, the background music factor can only be of relevance in those conditions in which the participants actually saw a particular video (and not the other video). The main effect alone therefore is not helpful, as it includes the unseen video. We therefore analyzed the seen video separately. To do so, we compared the effect of the background music (absent vs. present) for participants who answered the knee questions after seeing the knee video with a *t*-test for independent samples. This test confirmed that the background music had a detrimental effect on answering learning-congruent questions in the knee video, $t(87.81)=2.38, p=.025, d=0.48$.

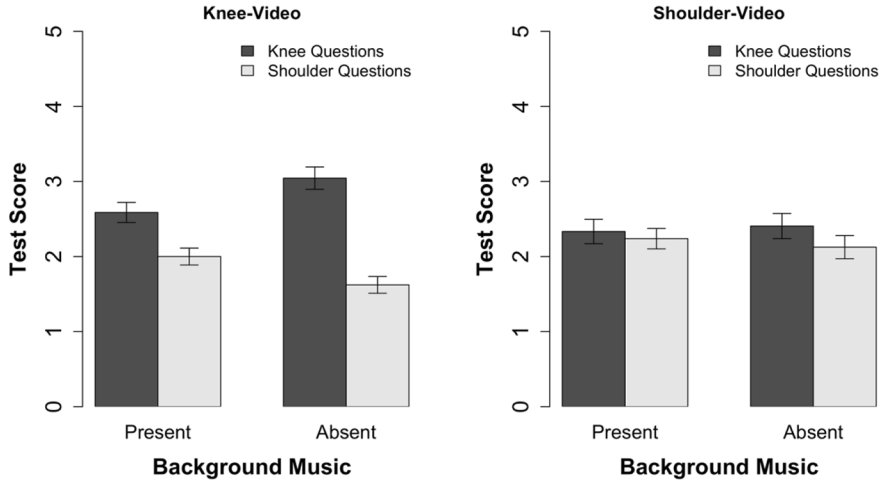


Fig. 2 Results of the knowledge tests. The content of the video as well as the presence of background music was manipulated between participants. The congruency between the content of the video and the questions was manipulated within participants (i.e., each participant saw only one video but answered both types of questions). Error bars indicate the SEM

For the shoulder questions (i.e., dark grey bars in Fig. 2), the follow-up analysis revealed a main effect of video content, $F(1, 161)=8.02, p=.005, \eta_G^2=0.05$, indicating that participants who attended to the shoulder video were more accurate in responding to the shoulder questions than participants who attended to the knee video. Further, the main effect of background music reached significance with better test performance with than without background music, $F(1, 161)=4.22, p=.042, \eta_G^2=0.03$, but there was no interaction between both variables, $F(1, 161)=1.08, p=.301, \eta_G^2<0.01$. At first sight, the main effect of the background music seems to indicate a beneficial effect of the background music. This main effect, however, includes the unseen video for which the background music cannot be of relevance. Therefore, we again analyzed the seen video separately. We compared the effect of the background music (absent vs. present) for participants who answered the shoulder questions after seeing the shoulder video with a t -test for independent samples. This test revealed no evidence for an impact of the background music on answering the congruent questions in the shoulder video, $t(67.17)=0.55, p=.583, d=0.13$.

Prediction Score

The analysis of the predicted test score (see Fig. 3) as the dependent variable was identical to the analysis of the actual test score. The three-way interaction between all variables was not significant, $F(1, 165)=2.17, p=.143, \eta_G^2<0.01$. However, we observed a two-way interaction between the content of the video and the congruency, $F(1, 165)=21.02, p<.001, \eta_G^2=0.03$. As confirmed by post-hoc t -tests for dependent samples, this interaction emerged from the presence of an effect of congruency for participants who attended to the knee video, $t(89)=8.01, p<.001, d_z=0.84$, but the absence of such an effect of congruency for participants who attended to the shoulder video, $t(78)=1.52, p=.132, d_z=0.17$. The remain-

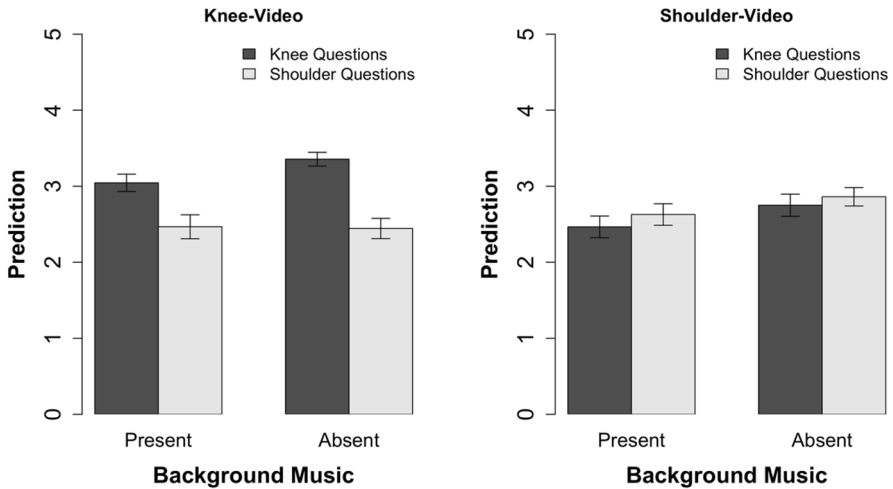


Fig. 3 Predicted test score (0–5) prior to the knowledge test. The content of the video as well as the presence of background music was manipulated between participants. The congruency between the content of the video and the questions was manipulated within participants. Error bars indicate the SEM

ing two-way interactions were not significant, all $F(1, 165) < 1.40$, all $p > .238$, all $\eta_G^2 < 0.01$. With regard to main effects, there was an effect of congruency with higher predictions for learning-congruent than incongruent questions, $F(1, 165) = 50.32$, $p < .001$, $\eta_G^2 = 0.07$, a marginal effect of background music with numerically lower predictions with than without background music, $F(1, 165) = 2.93$, $p = .089$, $\eta_G^2 = 0.01$, but no effect of the video content, $F(1, 165) = 1.78$, $p = .184$, $\eta_G^2 < 0.01$.

Questionnaire

We analyzed each scale/item in the questionnaire with a separate ANOVA for between-subject designs. Please note that most of the measurements in our questionnaire consist of a single item. The only exception are the scales for intrinsic (two items; observed Cronbach’s $\alpha = 0.67$), germane (three items; observed Cronbach’s $\alpha = 0.50$), and extraneous (three items; observed Cronbach’s $\alpha = 0.79$) cognitive load. For each analysis, we excluded those participants who did not respond to the item or to at least one of the items of the scale (we refrained from replacing missing values in scales, as these scales consisted of 2–3 items only).

The number of participants in each analysis, the mean values and standard deviation of each condition, as well as the corresponding test statistics are summarized in Table 1. For the sake of brevity, we address only those variables that may explain the relationship of background music and learning outcomes from a theoretical perspective.

The first relevant observation addresses experienced mental effort for which we observed an interaction between the video content and the background music. This interaction indicated that the addition of background music increased the invested mental effort only in the shoulder video, $t(56.04)$, $p = .007$, $d = 0.63$, but not the knee video, $t(88.94)$, $p = .400$, $d = 0.18$. With regard to the perceived demand characteristics, however, we observed a main

effect of background music, suggesting that videos with background music were experienced as less demanding than videos without background music.

A second relevant observation arises from the cognitive load scales. In contrast to mental effort, the background music had no effect on cognitive load. While we observed main effects of the video content for each of the three cognitive load scales, there were no main effects of the background music nor any interactions. With regard to the effects of the video content, the intrinsic and extraneous cognitive load were higher for the shoulder than the knee video, whereas the germane cognitive load was higher for the knee than the shoulder video.

The third and final relevant observation addresses the aesthetic experience of the participants. Here, the auditory track was not rated differently across all conditions, and the participants in the conditions without background music did not consider breaks between the spoken instructions to be more disturbing than the participants in the condition with background music.

Discussion

In this field experiment, we investigated whether instructional videos for rehearsal within a formal medical training improve learning (rehearsal hypothesis) and how the presence of background music affects the effectiveness of the instructional videos in imparting knowledge (background music hypothesis). Further, we aimed at investigating whether the students of the curriculum (who have prior experience with the learning domain medicine) appropriately predict their test performance relative to the effectiveness of the different versions of the instructional videos. In order to probe the generality of our effects, we exploratorily varied the difficulty the two different instructional videos.

Rehearsal with Instructional Videos

The broadest research question in the context of our project was whether implementing the instructional videos as rehearsal units into the curriculum improves test performance. We tested this question with the rehearsal hypothesis, which predicted such a beneficial effect. We used the answers to the questions addressing the unseen video as control conditions for the effectiveness of the instructional videos within participants (i.e., congruent vs. incongruent questions) as well as between participants (i.e., participants who saw the video addressing a particular set of questions vs. the other video).

With regard to the rehearsal hypothesis, our results confirm that both videos were effective in improving the participants' test performance. Most clearly, this is visible in the between-participant comparison, indicating that participants who had watched a particular video during rehearsal achieved higher test scores in the test for the content of this video than participants who had watched the other video. The within-participants comparisons were a bit more ambiguous. Whereas participants who had seen the knee video for rehearsal achieved a higher test score for the questions addressing the knee video than the questions addressing the shoulder video, this pattern did not emerge vice versa; that is, the participants who had seen the shoulder video did not achieve a higher test score for the questions addressing the shoulder video than those addressing the knee video. Please note, however,

Table 1 Summary of the analysis of the questionnaire

Item/Scale	Condition			Statistics		
	Knee BM+	Knee BM-	Shoulder BM-	ME Video Content	ME Background Music	Interaction
N	M (SD)	M (SD)	M (SD)	M (SD)	M (SD)	M (SD)
Joy of Learning						
Joy	170 4.15 (1.23)	4.02 (1.24)	3.18 (1.40)	3.50 (1.47)	$F(1, 166)=13.81, p<.001$	$F(1, 166)=1.19, p=.278$
Professionalism						
Video	173 5.85 (1.10)	5.61 (1.13)	5.43 (1.28)	5.90 (1.14)	$F(1, 169)<.1$	$F(1, 169)=4.77, p=.030$
Auditory Track	173 5.89 (1.10)	5.70 (1.15)	5.72 (1.14)	5.82 (1.17)	$F(1, 169)<.1$	$F(1, 169)<.1$
Visual Track	173 5.63 (1.06)	5.16 (1.31)	5.00 (1.54)	5.56 (1.21)	$F(1, 169)<.1$	$F(1, 169)=6.92, p=.009$
Aesthetic Experience						
Video	172 5.02 (1.36)	4.86 (1.27)	4.11 (1.37)	4.45 (1.41)	$F(1, 168)=10.70, p=.001$	$F(1, 168)=1.42, p=.235$
Disturbing Breaks	172 2.26 (1.57)	2.41 (2.02)	2.07 (1.50)	2.53 (1.64)	$F(1, 168)<.1$	$F(1, 168)<.1$
Auditory Track	170 4.98 (1.48)	5.16 (1.15)	5.00 (1.41)	5.25 (1.34)	$F(1, 166)<.1$	$F(1, 166)<.1$
Experienced Effort						
Mental Effort	174 4.00 (1.46)	4.27 (1.56)	5.58 (0.82)	4.79 (1.58)	$F(1, 170)=25.93, p<.001$	$F(1, 170)=6.23, p=.014$
Demand Characteristics	174 2.57 (1.17)	3.20 (1.29)	3.30 (1.39)	3.51 (1.59)	$F(1, 170)=6.55, p=.011$	$F(1, 170)=1.01, p=.316$
Cognitive Load						
Intrinsic Load	174 3.77 (1.18)	3.68 (1.16)	4.45 (1.32)	4.14 (1.50)	$F(1, 170)=8.76, p=.004$	$F(1, 170)<.1$
Germane Load	174 5.04 (1.06)	5.30 (0.72)	4.72 (1.00)	4.85 (1.11)	$F(1, 170)=6.62, p=.011$	$F(1, 170)<.1$
Extraneous Load	174 3.03 (1.21)	3.13 (1.34)	4.37 (1.48)	3.99 (1.71)	$F(1, 170)=26.19, p<.001$	$F(1, 170)=1.20, p=.276$
Overall Assessment						
Likings	175 4.85 (1.27)	4.78 (1.44)	3.66 (1.40)	3.95 (1.59)	$F(1, 171)=22.38, p<.001$	$F(1, 171)<.1$
Further Units	175 4.70 (1.55)	5.00 (1.48)	3.86 (1.75)	4.26 (1.89)	$F(1, 171)=9.93, p=.002$	$F(1, 171)<.1$
Motivation	175 3.87 (1.61)	4.27 (1.44)	3.43 (1.59)	3.21 (1.73)	$F(1, 171)=9.38, p=.003$	$F(1, 171)=1.66, p=.200$

Note: BM+= video with background music; BM- = video without background music

that not only the shoulder video but also the shoulder topic in general was more complex (i.e., more examination techniques). Therefore, the lack of a within-subject effect for the participants who had watched the shoulder video likely reflects the complexity of the topic. In other words, as the shoulder topic was more complex, the initial learning session and the rehearsal video resulted in the same level of the test scores as the initial learning session alone for the knee content. In particular, given the unambiguous results of the comparison between participants, we therefore accept the rehearsal hypothesis; that is, the instructional videos were effective in communicating knowledge.

Our findings regarding the rehearsal hypothesis agree with the large body of research emphasizing the importance of rehearsal for learning (e.g., Craik & Watkins 1973; Woodward et al., 1973). For such a rehearsal, instructional videos appear to be of particular interest because they can be studied at any time independent of formal classroom schedules. Further, although instructional videos convey the same knowledge as the initial teaching (especially when teaching is aligned between them), they vary in several surface features from the initial teaching setup. Among other things, they vary with respect to the instructor, the presented examples, and the learning environment (e.g., classroom vs. home). Importantly, such variations and also comparable variations of the learning context have been demonstrated to foster long-term learning rather than just short-term retrievability (e.g., Smith et al., 1978).

Background music in instructional videos

Previous research has revealed conflicting findings on how background music affects learning with instructional videos, ranging from beneficial effects (e.g., Flood 2007; Rauscher et al., 1993; Lehmann et al., 2019), to hardly any effect (e.g., Kopiez et al., 2013), to detrimental effects (Boeckmann et al., 1990; Brosius, 1990). As most of the previous studies have investigated samples without prior experience with the learning content, we decided to reinvestigate the impact of the background music within the rehearsal videos in our study.

The observed effect of the background music differed between the two instructional videos. Whereas the presence of background music lowered the test score in the memory test for the knee video, no such detrimental effect emerged in the shoulder video. It seems noteworthy, however, that in none of the videos did we observe any evidence for a beneficial impact of the background music on learning. The overall pattern of our data therefore tends to agree with the conclusions of Boeckmann et al., (1990) and Brosius (1990) who deemed background music as harmful for learning with instructional videos. In contrast to the previous studies, however, we used the instructional videos for rehearsal rather than initial learning. Although it appears unlikely, it remains possible that the impact of background music differs between initial learning and rehearsal.

Beyond the objective learning outcome, our questionnaire allowed us to explore some of the potential cognitive processes that might be responsible for the detrimental effect of the background music. In previous work, such detrimental effects of background music have been conceptualized as a variant of the seductive details effect (Park et al., 2011). In a systematic review, Rey (2012) identified four potential mechanisms explaining the detrimental effects of seductive details on learning. First, seductive details might overload working memory. Second, they might distract attention away from the relevant materials. Third, they might interfere with the schema to be learnt and, fourth, they might disrupt the coherence

of the learning materials. With regard to our present project, only a potential overload of the working memory appears to be a plausible candidate for turning the background music into a seductive detail. However, with regard to the measures of cognitive load, we did not observe any differences between the videos with and without background music. In particular, the measure of extraneous cognitive load (i.e., learning irrelevant cognitive load) was not influenced by the presence of background music (also see Brünken et al., 2004). Our data therefore does not support the assumption that background music acts as a seductive detail in our instructional videos and thus decreases learning. Of course, however, it also remains possible that a potential overload induced by background music could be rather implicit and therefore might not be reflected in subjective measures such as the cognitive load scales used in this experiment. Therefore, future experiments could use implicit and more objective measures of cognitive load, such as pupil dilation (Lee et al., 2020).

An alternative cognitive process for explaining the detrimental effect of the background music arises from the lowered perceived demand characteristics of the videos when background music was present rather than absent. Such a reduction in the perceived demand characteristics has been identified to potentially lower learning outcomes (Salomon, 1984; for related recent findings see Ryffel & Wirth, 2020; Uzun & Yilderim, 2018). However, the lowered perceived demand characteristics emerged as a main effect of the background music, whereas learning was reduced only in the knee video. Therefore, there are likely further factors influencing the detrimental effect of the background music on learning. As both videos and their topics varied in their formal difficulty in terms of the number of examination techniques, difficulty appears to be a straightforward candidate for such a moderating factor. The formal difficulty also matched the subjective ratings; for instance, the shoulder video was experienced as more demanding in terms of perceived demand characteristics as well as invested mental effort. The ratings regarding cognitive load also suggest that the shoulder video (or its content) was less suitable for learning (lower germane cognitive load but higher intrinsic and extraneous cognitive load) than the knee video. A potential mechanism that deserves attention in future research is therefore that a high difficulty of the video content motivates participants to invest a maximum of mental effort, thus diminishing the detrimental effect of the background music.

Prediction of Test Performance

Beyond actual test scores, we were interested in whether the pattern of perceived learning (i.e., the predicted test outcome) matched the test outcome. In other words, we were interested in whether the predicted test scores were sensitive to the same manipulations as the actual test scores. The main motivation for studying this was the remarkable mismatch between both scores in previous work, which seemed to arise from a tendency to misattribute the impression of fluency to successful learning (Carpenter et al., 2013; Kornell & Bjork, 2008; Ryffel & Wirth, 2020; Toftness et al., 2018). We investigated this question with our *learning prediction hypothesis*.

We observed that the pattern of predicted test scores appeared to match the pattern of the actual test scores. In particular, we observed the same pattern of results with regard to rehearsal with the knee versus the shoulder video as in the actual knowledge test. Whereas participants who had watched the knee video for rehearsal expected to be more accurate with regard to knee questions than to the shoulder questions, no such effect emerged within

participants who had watched the shoulder video for rehearsal. The (partially) detrimental effect of the background music on learning did not fully emerge in the prediction scores, but the corresponding main effect of the background music trended toward significance. Overall, this pattern is consistent with the idea that participants who have experience with a particular learning domain (such as medical training) can (roughly) predict their success of learning with a particular instructional video (i.e., they are less susceptible to illusions of learning; see Kardas & O'Brien 2018).

Although the effect of background music on the predicted test score only tended toward significance, the numerical direction of this effect contrasts with the direction one would expect from considerations about fluency. For instance, Ryffel and Wirth (2020) reported that the presence of background music increases the perception of fluency during learning. In agreement with these findings, the perceived demand characteristics of the variants of our instructional videos with background music were rated lower than those without background music. Based on the previous findings, one would now expect an increase in the predicted learning outcome due to the misattribution of fluency toward learning. However, we observed the opposite pattern: that is, lower predicted test scores after instructional videos with background music rather than without (at least numerically). This surprising finding suggests that the participants in our study were able to predict more robust learning under conditions with less experienced fluency (e.g., Diemand-Yauman et al., 2011; but see Kühl et al., 2014). In return, this suggests that our participants were less susceptible to the misattribution of experienced fluency to learning as the participants in previous samples (e.g., Carpenter et al., 2013; Kornell & Bjork, 2008).

The most striking difference between the prior studies and our field experiment was that we recruited a sample that had prior experience with the overall content of learning. As medical students (in the 9th semester of their studies) are experts when it comes to studying medical procedures, we think that they potentially might be able to distinguish between fluency and learning. Nevertheless, the background music was harmful for learning with the knee video. It is an interesting question for future research to investigate whether the accurate self-evaluation of learning is due to the participants' detection of worse comprehension already during the learning phase or whether these predictions are generated in the moment that participants are asked to predict their learning outcomes. Further, we think that studying the impact of learning experience with a particular content of learning should be of major relevance for future research addressing the interplay between perceived fluency and learning. We consider it likely that the susceptibility to misattributions of fluency to learning acts as a heuristic when experience with the content of learning is lacking or low. In other words, the impact of perceived fluency should diminish with increasing expertise in a particular learning content.

Strengths, Limitations, and Outlook

A noteworthy strength of our study is that we conducted the experiment within an applied setting (i.e., a field experiment). Our participants were medical students who took part in the rehearsal unit with the instructional video as a part of a one-day internship. Therefore, our participants were not only highly motivated to actually learn the studied information but also had prior experience with the general content (e.g., learning medical procedures), but not with the specific learning materials. In the prediction scores, this prior experience is

the most straightforward candidate mechanism that might have immunized our participants against illusions of knowing that have been observed in prior work following misattributions of perceptual fluency to successful learning (Carpenter et al., 2013; Ryffel & Wirth, 2020). To our knowledge, the impact of experience with the learning content on the illusion of knowing has not yet been addressed in research. However, this relationship needs to be studied in future work, as it might reflect an important moderating factor for the success of instructional videos across various samples.

Of course, our study also has limitations, some of which arise from the applied setting. As our field experiment was implemented into a regular curriculum, there were temporal constraints which did not allow for extensive testing. For instance, we could not run a full test on prior knowledge. However, given the rather specific topic within a rather strict curriculum, systematic differences in prior knowledge before the internship seem unlikely. More importantly, the videos that we implemented in our experimental manipulations served as rehearsal for the course contents in the curriculum, so that we can assume that any potential differences in prior knowledge before the internship were leveled out in the course input phase that was conducted before our experimental manipulations. In order to control that the participants actually learn from the rehearsal with our instructional videos, we relied on between-participants comparisons for seen and unseen videos (i.e., participants who had seen a particular video were more accurate in answering the corresponding knowledge questions than participants who had seen the other video). More generally, the knowledge tests had to be rather short due to the temporal restrictions. Longer tests might have revealed clearer results with regard to the shoulder video (and therefore should be embedded in future research).

Further, we were only able to briefly address the cognitive processes behind the impact of background music on learning but had to refrain from studying other potential impacts of the background music, which has been discussed in previous work such as indirect effect via changes in mood or arousal (Boltz et al., 1991; Husain et al., 2002; Thompson et al., 2011; Lehmann et al., 2019). For instance, Bellier et al. (2020) observed that background music reduced anxiety in dissection classes. Please note, however, that we did not observe an effect of the background music on the motivation to learn with the video or the overall assessment of the video. This suggests the absence of such indirect effects on the motivation to learn. Again, a key difference might be in the selected sample. Whereas participants who do not need to acquire knowledge in order to master a university course may be more susceptible to moderations of the motivation to learn, a highly motivated sample recruited within a teaching curriculum might be less susceptible to the same cues.

One further limitation of our study is that we were not able to test the effect of the rehearsal with the instructional video against other forms of rehearsal as well as longer retention intervals. Our study showed that our participants had more knowledge about the examination techniques of either the knee or the shoulder joint when they rehearsed the content with an instructional video for the respective joint rather than the other joint. However, future research should investigate more systematically how the effect of rehearsal with instructional videos relate to other forms of rehearsal. In addition, such a study might investigate the long-term impact of rehearsal.

We consider our main finding that rehearsal with instructional videos improves learning as an encouraging starting point for a future research program that investigates further how instructional videos could be implemented into a curriculum to optimize the learning out-

come. With regard to this implementation, the most intriguing research question is whether it makes more sense to use the instructional video for rehearsal or for preparation (i.e., flipped classroom; van Alten et al., 2019). Further, because background music was identified as one feature of the more popular instructional videos on YouTube (ten Hove & van der Meij, 2015), it would be another interesting pathway for future research whether other characteristics that make videos popular (e.g., fast speed of narration, production quality) indeed facilitate or rather hamper learning.

Conclusions

Our field experiment confirmed the effectiveness of instructional videos as a tool for rehearsal within a sample of medical students in their regular curriculum. With regard to the design of such instructional videos, we did not observe any evidence in favor of adding background music to the videos. In contrast, background music was even harmful for rehearsing in one of the two videos under investigation. The most likely process behind this is that the background music lowered the perceived demand characteristics of that video. Based on these results, we would recommend to present instructional videos without background music (although we admit that more research is necessary to fully understand how and why background music affects learning with videos). However, our study also revealed that the difficulty of the content of the video might alter the impact of features such as background music. In general, more research is necessary to systematically explore the optimal implementation of an instructional video into a formal curriculum of experienced learners as well as to establish design principles for such videos that focus on optimizing learning outcomes.

Authors' contributions All authors contributed to the study design. HSM, MM, and CS designed the psychological questionnaire. AM designed the medical knowledge test. AM collected the data. HSM analyzed the data. HSM drafted the manuscript. All authors provided critical revisions to the manuscript.

Funding No funding was involved in this project.

Data Availability Materials, data, and scripts are available at <https://osf.io/9u5wr/>.

Declarations

Conflicts of interest/Competing interests None of the authors have a no conflict of interest or competing interests.

References

- Ayres, P., & Paas, F. (2007). Can the cognitive load approach make instructional animations more effective? *Applied Cognitive Psychology, 21*, 811–820
- Azer, S. A., AlKhawajah, N. N., & Alshamlan, Y. A. (2022). Critical evaluation of YouTube videos on colostomy and ileostomy: Can these videos be used as learning resources? *Patient Education and Counseling, 105*, 383–389. <https://doi.org/10.1016/j.pec.2021.05.023>
- Biard, N., Cojean, S., & Jamet, E. (2018). Effects of segmentation and pacing on procedural learning by video. *Computers in Human Behavior, 89*, 411–417

- Bjork, R. A., & Bjork, E. L. (1992). A new theory of disuse and an old theory of stimulus fluctuation. *From Learning Processes to Cognitive Processes: Essays in Honor of William K Estes*, 2, 35–67
- Boeckmann, K., Nessmann, K., Petermandl, M., & Stückler, H. (1990). On the influence of background music on recall and appraisal in educational films. *Educational Media International*, 27, 172–180
- Boltz, M., Schulkind, M., & Kantra, S. (1991). Effects of background music on the remembering of filmed events. *Memory & Cognition*, 19, 593–606
- Boucheix, J. M., Gauthier, P., Fontaine, J. B., & Jaffeux, S. (2018). Mixed camera viewpoints improve learning medical hand procedure from video in nurse training? *Computers in Human Behavior*, 89, 418–429
- Brosius, H. B. (1990). Bewertung gut, Behalten schlecht: Die Wirkung von Musik in Informationsfilmen [Good evaluation, bad memory: The effect of music in information films]. *Medienpsychologie*, 2, 44–55
- Brünken, R., Plass, J. L., & Leutner, D. (2004). Assessment of cognitive load in multimedia learning with dual-task methodology: Auditory load and modality effects. *Instructional Science*, 32, 115–132
- Carpenter, S. K., Wilford, M. M., Kornell, N., & Mullaney, K. M. (2013). Appearances can be deceiving: Instructor fluency increases perceptions of learning without increasing actual learning. *Psychonomic Bulletin & Review*, 20, 1350–1356
- Craik, F. I., & Watkins, M. J. (1973). The role of rehearsal in short-term memory. *Journal of Verbal Learning and Verbal Behavior*, 12, 599–607
- DeFleur, M. L., Davenport, L., Cronin, M., & DeFleur, M. (1992). Audience recall of news stories presented by newspaper, computer, television and radio. *Journalism Quarterly*, 69, 1010–1022
- de la Mora Velasco, E., & Hirumi, A. (2020). The effects of background music on learning: a systematic review of literature to guide future research and practice. *Educational Technology Research and Development*, 68, 2817–2837
- Diemand-Yauman, C., Oppenheimer, D. M., & Vaughan, E. B. (2011). Fortune favors the bold (and the italicized): Effects of disfluency on educational outcomes. *Cognition*, 118, 111–115
- Eschrich, S., Münte, T. F., & Altenmüller, E. O. (2008). Unforgettable film music: the role of emotion in episodic long-term memory for music. *BMC Neuroscience*, 9, 1–7
- Fiorella, L., & Mayer, R. E. (2018). What works and doesn't work with instructional video. *Computers in Human Behavior*, 89, 465–470
- Flood, M. (2007). The relationships among background music in the classroom, attitudes toward nursing research, and academic achievement. *Journal of Best Practices in Health Professions Diversity*, 117–128
- Furnham, A., Gunter, B., & Green, A. (1990). Remembering science: The recall of factual information as a function of the presentation mode. *Applied Cognitive Psychology*, 4, 203–212
- Gunter, B., Furnham, A. F., & Gietson, G. (1984). Memory for the news as a function of the channel of communication. *Human Learning: Journal of Practical Research & Applications*, 3, 265–271
- Höffler, T. N., & Leutner, D. (2007). Instructional animation versus static pictures: A meta-analysis. *Learning and Instruction*, 17, 722–738
- Hoogerheide, V., van Wermeskerken, M., Loyens, S. M., & van Gog, T. (2016). Learning from video modeling examples: Content kept equal, adults are more effective models than peers. *Learning and Instruction*, 44, 22–30
- Husain, G., Thompson, W. F., & Schellenberg, E. G. (2002). Effects of musical tempo and mode on arousal mood and spatial abilities. *Music Perception*, 20, 151–171
- Kardas, M., & O'Brien, E. (2018). Easier seen than done: Merely watching others perform can foster an illusion of skill acquisition. *Psychological Science*, 29, 521–536
- Kay, R. H. (2012). Exploring the use of video podcasts in education: A comprehensive review of the literature. *Computers in Human Behavior*, 28, 820–831
- Kay, R., & Kletschin, I. (2012). Evaluating the use of problem-based video podcasts to teach mathematics in higher education. *Computers & Education*, 59, 619–627
- Klepsch, M., Schmitz, F., & Seufert, T. (2017). Development and validation of two instruments measuring intrinsic, extraneous, and germane cognitive load. *Frontiers in Psychology*, 8, 1997
- Kopiez, R., Platz, F., & Wolf, A. (2013). The overrated power of background music in television news magazines: A replication of Brosius' 1990 study. *Musicae Scientiae*, 17, 309–331
- Kornell, N., & Bjork, R. A. (2008). Learning concepts and categories: Is spacing the “enemy of induction”? *Psychological Science*, 19, 585–592
- Kühl, T., Eitel, A., Scheiter, K., & Gerjets, P. (2014). A call for an unbiased search for moderators in disfluency research: reply to Oppenheimer and Alter (2014). *Applied Cognitive Psychology*, 28, 805–806
- Lee, J. Y., Donkers, J., Jarodzka, H., Sellenraad, G., & van Merriënboer, J. (2020). Different effects of pausing on cognitive load in a medical simulation game. *Computers in Human Behavior*, 110, 106385
- Lehmann, J. A. M., Hamm, V., & Seufert, T. (2019). The influence of background music on learners with varying extraversion: Seductive detail or beneficial effect? *Applied Cognitive Psychology*, 33, 85–94
- Lehmann, J. A., & Seufert, T. (2017). The influence of background music on learning in the light of different theoretical perspectives and the role of working memory capacity. *Frontiers in Psychology*, 8, 1902

- Mayer, R. E. (2001). *Multimedia Learning*. Cambridge, UK: Cambridge University Press
- Meder, A., Stefanescu, C., Ateschrang, A., Fröhlich, S., Obertacke, U., Schulz, A. P. ... Sterz, J. (2021a). Evidenzbasiertes Lehrvideo für Untersuchungstechniken am Kniegelenk. *Zeitschrift für Orthopädie und Unfallchirurgie*, *159*, 454–457
- Meder, A., Stefanescu, C., Ateschrang, A., Fröhlich, S., Obertacke, U., Schulz, A. P. ... Ruesseler, M. (2021b). Evidenzbasiertes Lehrvideo für Untersuchungstechniken am Schultergelenk. *Zeitschrift für Orthopädie und Unfallchirurgie*, *159*, 332–335
- Meitz, T. G., Meyerhoff, H. S., & Huff, M. (2020). Event related message processing: Perceiving and remembering changes in films with and without soundtrack. *Media Psychology*, *23*, 733–763
- Merkt, M., Lux, S., Hoogerheide, V., van Gog, T., & Schwan, S. (2020). A change of scenery: Does the setting of an instructional video affect learning? *Journal of Educational Psychology*, *112*, 1273–1283
- Meyerhoff, H. S., & Huff, M. (2016). Semantic congruency but not temporal synchrony enhances long-term memory performance for audio-visual scenes. *Memory & Cognition*, *44*, 390–402
- Olejnik, S., & Algina, J. (2003). Generalized eta and omega squared statistics: measures of effect size for some common research designs. *Psychological Methods*, *8*, 434–447
- Paas, F. G. (1992). Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive-load approach. *Journal of Educational Psychology*, *84*, 429–434
- Park, B., Moreno, R., Seufert, T., & Brünken, R. (2011). Does cognitive load moderate the seductive details effect? A multimedia study. *Computers in Human Behavior*, *27*, 5–10
- Ploetzner, R., Berney, S., & Bétrancourt, M. (2020). A review of learning demands in instructional animations: The educational effectiveness of animations unfolds if the features of change need to be learned. *Journal of Computer Assisted Learning*, *36*, 838–860
- Rauscher, F. H., Shaw, G. L., & Ky, K. N. (1993). Music and spatial task performance. *Nature*, *365*, 611
- Rey, G. (2012). A review of research and a meta-analysis of the seductive detail effect. *Educational Research Review*, *7*, 216–237
- Ruesseler, M., Weber, R., Braunbeck, A., Flaig, W., Lehrteam des Zenrum Chirurgie, Marzi, I., & Walcher, F. (2010). Training praktischer Fertigkeiten in der Chirurgie – Ein Ausbildungskonzept für Studierende [Training of practical clinical skills in surgery—a training concept for medical students]. *Zentralblatt Fur Chirurgie*, *135*(3), 249–256
- Ryffel, F. A., & Wirth, W. (2020). How perceived processing fluency influences the illusion of knowing in learning from TV reports. *Journal of Media Psychology: Theories, Methods, and Applications*, *32*(1), 2–13. <https://doi.org/10.1027/1864-1105/a000250>
- Salomon, G. (1984). Television is” easy” and print is” tough”: The differential investment of mental effort in learning as a function of perceptions and attributions. *Journal of Educational Psychology*, *76*, 647–658
- Schmider, E., Ziegler, M., Danay, E., Beyer, L., & Bühner, M. (2010). Is it really robust? *Methodology*, *6*, 147–151
- Schwan, S., & Riempp, R. (2004). The cognitive benefits of interactive videos: Learning to tie nautical knots. *Learning and Instruction*, *14*, 293–305
- Smith, S. M., Glenberg, A., & Bjork, R. A. (1978). Environmental context and human memory. *Memory & Cognition*, *6*, 342–353
- Smith, T. J., & Martin-Portugues Santacreu, J. Y. (2017). Match-action: the role of motion and audio in creating global change blindness in film. *Media Psychology*, *20*, 317–348
- Stadler, M., Sailer, M., & Fischer, F. (2021). Knowledge as a formative construct: A good alpha is not always better. *New Ideas in Psychology*, *60*, 100832
- Sweller, J., Van Merriënboer, J. J., & Paas, F. G. (1998). Cognitive architecture and instructional design. *Educational Psychology Review*, *10*, 251–296
- ten Hove, P., & van der Meij, H. (2015). Like it or not. What characterizes YouTube’s more popular instructional videos? *Technical Communication*, *62*, 48–62
- Thompson, W. F., Schellenberg, E. G., & Letnic, A. K. (2011). Fast and loud background music disrupts reading comprehension. *Psychology of Music*, *40*, 700–708
- Toftness, A. R., Carpenter, S. K., Geller, J., Lauber, S., Johnson, M., & Armstrong, P. I. (2018). Instructor fluency leads to higher confidence in learning, but not better learning. *Metacognition and Learning*, *13*, 1–14
- Uzun, A. M., & Yıldırım, Z. (2018). Exploring the effect of using different levels of emotional design features in multimedia science learning. *Computers & Education*, *119*, 112–128
- van Alten, D. C., Phielix, C., Janssen, J., & Kester, L. (2019). Effects of flipping the classroom on learning outcomes and satisfaction: A meta-analysis. *Educational Research Review*, *28*, 100281
- Woodward, A. E. Jr., Bjork, R. A., & Jongeward, R. H. Jr. (1973). Recall and recognition as a function of primary rehearsal. *Journal of Verbal Learning and Verbal Behavior*, *12*, 608–617

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.