Analysing the Characteristics of YouTube's Popular Instructional Videos for Software Training and Proposing Guidelines for their Design

Constanze Höpfner
M.Sc. Thesis Educational Science and Technology
University of Twente
Faculty of Behavioural, Management and Social Sciences
Department of Instructional Technology

Examination Committee

First supervisor:
Dr. A.M. van Dijk
a.m.vandijk@utwente.nl

Second supervisor:
Dr. H. van der Meij
h.vandermeij@utwente.nl

13 November 2020
Enschede, The Netherlands

Author's note

Please note that this thesis cannot be published in the UT depository because it will be published as a journal article. A request for confidentiality has been filed.

CONFIDENTIAL

Table of Contents

Acknowledgements

Writing this thesis has been a challenging, fruitful and rewarding experience.
First of all, I would like to thank my supervisors Dr. Alieke van Dijk and Dr. Hans van der Meij for the feedback, support and their open and friendly manner of coaching.
In addition, I would like to thank my second coder for his time and input.

With this thesis, my time as a Master student officially comes to an end.
It has been a journey of immense personal and professinal growth- thanks to the courses, working for Komma and most of all, thanks to all the people that I met along the way.
I am very grateful to have had this opportunity to learn and grow.
Last but not least, all of this would have not been the same without the love and support from my friends and family.

To all of you - thank you.

Abstract

YouTube videos have become an immensely popular tool to convey how-to instructions for software training. Such videos can be an effective educational tool when instructional design principles are applied. Yet, so far few studies have analysed if they are applied in YouTube software tutorials and if this affects video popularity. To close this gap, this study addresses the following questions: What are the instructional design characteristics of YouTube's popular video software tutorials? How do these characteristics affect video popularity? And finally, which design guidelines can be proposed to support the theoretical effectiveness of these videos, and their popularity? To answer these questions, a codebook is developed, which is based design priniciples known from literature such as the multimedia principles. Subsequently, this codebook is used to determine the characteristics of 45 popular video software tutorials published on YouTube, and statistical analyses are run to determine the relationship between video characteristics and video popularity. The results show that many multimedia principles are applied, and that numerous principles are closey intertwined with the narrative.This highlights the decisive role of the narrative for the effectiveness of instructional video. Further, only two of the assessed video characteristics significantly influence video popularity: showing the narrator's face positively influences popularity, while providing more procedural information negatively influences popularity. Taking the study outcomes, a literature review and YouTube's platform-specific features into account, guidelines for the design of theoretically effective video software tutorials are proposed; including specific advice on how write and pace a narrative, and on how to employ YouTube's chapter.

Keywords: instructional video, YouTube, procedural support, software tutorials, multimedia learning.

Introduction

Information seekers and learners around the world turn to YouTube videos to educate themselves. YouTube is a social media platform and one of the most widely used video platforms for learning (Khan, 2017; Haridakis & Hanson, 2009). It has an enormous user base of approximately 1,9 billion monthly logged-in users (Clement, 2019), and about 30 million daily worldwide visitors. While YouTube offers a wide range of videos and is well-known for offering videos for entertainment, a recent Google survey revealed that, in 2017, the YouTube content category that received most attention by viewers was how-to video (Think with Google, 2017). Users state to make use of the platforms' videos in order solve a problem (54%), to fix something in their home or car (65%) or to improve their school or jobs skills (37%), to name some examples (O'Neil-Hart, 2018).

Despite its widespread use for learning purposes, YouTube remains an under-researched platform compared to other social media such as Facebook or Twitter (Khan, 2017; Snelson, Rice & Wyzard, 2012). Snelson et al. (2012) determined and prioritised categories for research on YouTube and other video-sharing technologies, and education was ranked as the second highest priority category. In particular, the identification of instructional design elements of effective online videos and success factors of educational videos were earmarked (Snelson et al., 2012). Nevertheless, since then only a small number of studies dealt with these topics.

Swarts (2012), and Morain and Swarts (2012) conducted a qualitative study on 46 of YouTube's instructional videos for software training. The authors divided the videos into three groups, popular, average and poorly rated videos, based on YouTube former one to five-star user rating system and analysed how much time videos spent on introduction, actions steps and conclusions. The video analysis included features related to facilitating access (e.g. headings), viewability (e.g. video and audio quality), timing (e.g. narration speed). Further, they assessed how accurate, complete and pertinent presented information was with regard to the instructional goal, and to what extent the narrator radiated self-confidence and supported the viewers' self-efficacy and engagement. They found that popular videos, in contrast to average and poor ones, spent more time introducing and framing goals and actions steps. Further, popular videos concentrated on doing and explaining steps, and included a message that is "easy to locate and access, easy to understand and utilise, and is engaging and reassuring" (Swarts, 2012, p.195). However, as noted by Ten Hove and van der Meij (2015, p. 50), "most of the features assessed by Swarts hinged on interpretation" (p.50), which makes it difficult apply their findings.

From his research, Swarts (2012) derived best practices for video design, and their presence was investigated in a recent study by Chong (2018) about YouTube's top ten most-viewed make-up and hair tutorials. Chong (2018) found several similarities to Swarts' findings: the majority of videos contained an introduction that stated a clear objective, timed audio and video, and showed how to successfully achieve a task in the first attempt and focused on conveying verbal instructions. In contrast, there was a number of video features that clearly deviated from Swarts' best practices (2012). In Chong's (2018) study, a minority of videos appeared to have a well-rehearsed script; video demonstrated high quality audio and video; was edited to achieve a flawless video; and also, none of the videos included reassurances to promote the viewers' self-efficacy - a key recommendation of Swarts. In addition, Chong assessed the videos with regards to presence of assertions of credibility (Mackiewicz, 2010) the vast majority of narrators did not make any explicit assertions of credibility with regards to their product-specific experience, familiarity with related or relevant products or their role as a make-up artist. Chong (2018) concludes that in the case of beauty tutorials, other factors than those proposed in Swarts (2012) best practices appear to drive video popularity, and suggests that features such as including personal narratives and humour may contribute to video popularity. Unfortunately, like the features proposed by Swarts' (2012), Chong's (2018) feature are subordinate to interpretation.

Ten Hove and van der Meij (2015) analysed quantitative, objectively determinable features of popular YouTube videos that teach factual and conceptual information. In this study, video popularity was measured based on a formula that included viewer ratings (such as views, likes, dislikes). They identified that popular videos differ significantly from unpopular or average videos. Typically, popular videos were characterised by higher video resolution, frequent presence of static pictures and of a combination of static and dynamic pictures, short on-screen text, subtitling in different languages, background music, less background noise and a faster speaking rate (Ten Hove and van der Meij, 2015).

Shoufan evaluated YouTube videos on dealing with "academic topics'' (Shoufan, 2017, p.128) without further specifying which type of knowledge they teach. Similar to Ten Hove and van der Meij (2015), Shoufan and Mohamed (2017) found that videos with higher resolution and speaking rate and a native English speaker received more positive user ratings; moreover, video that showed explanations on paper or whiteboard and videos that have more than one production style received more likes. Further, Shoufan (2019) examined YouTube videos about digital logic design with regards to general video features (e.g. production style and length) and the presence or absence of features described in Mayer and Moreno's multimedia principles (2003). These principles suggest how to optimise multimedia content presentation, that is the combination of words and pictures such as in video, to facilitate learning. Shoufan (2019) found that only four of the ten principles impacted viewer ratings and suggested that further research is needed to identify features of popular instructional videos.

While the previous work identified mainly general video characteristics, I am curious to take the video analysis a step further and identify which specific, objectively appraisable instructional design principles are present in YouTube's popular instructional videos. Instructional design principles suggest how to design information in a manner that supports cognitive processing and improves learning. To give an example, the signaling principle (Mayer & Moreno, 2003) holds that learning is enhanced when the material includes cues that guide the learner's attention to relevant elements. As described by van der Meij & van der Meij (2013), signaling can be applied by adding cues such as coloured frames, arrows or dotted lines to suitable parts of the video, but these are just a few examples. I am interested in how instructional design principles are employed in YouTube videos and if any particular type relates to video popularity, e.g. if the use of signaling in form of arrows leads to higher video popularity than using coloured frames. As pointed out by Garrett (2019), it is important to find out how learners use educational videos in authentic contexts, to be able to test instructional theories in these new contexts and evaluate their impact on learning adequately. YouTube appears to be a very relevant source for learners, so scrutinising the characteristics of popular instructional YouTube videos could be a starting point for further research into whether characteristics that make video popular are also effective for learning.

In particular, I will focus on YouTube's popular videos for software training, i.e. those that show how to execute software tasks. Learning such tasks requires learning procedural knowledge, which is defined as information on how to perform sequences of actions (Colman, 2008). Learning procedural knowledge is suggested to be more difficult than learning declarative knowledge (Anderson, 1995) and, so far, research on videos for software training is limited and inconclusive. While some studies found positive effects of video for teaching SPSS (Lloyd & Robertson, 2012) and Word (van der Meij & van der Meij, 2014) others found mixed or neutral results for video instructions for using Word (Alexander, 2013), SPSS (Brar & van der Meij, 2017) and Excel (Worlitz, Stabler, Peplowsky & Woll, 2016). Moreover, to the best of my knowledge, the only studies conducted on software tutorials in a YouTube context were performed in 2012 by Morain and Swarts, and Swarts (2012), so overall, research is still scant.

With the present study, I seek to determine the characteristics of video software tutorials on YouTube, how these relate to video popularity, and what recommendations can be proposed for the design of video software tutorials published on YouTube from the perspective of educational science. To achieve this, I will take the following steps. First, I will create a framework for the objective assessment of said videos based on proven and, for the YouTube context, relevant instructional design principles. To assure the reliability and objectivity of the framework, I will test and improve it until reaching a high inter-rater reliability. Second, I will identify a suitable formula for gauging video popularity from publicly available user appraisal data (i.e. number of views, likes and dislikes). Third, I will take a sample of 45 YouTube's most popular instructional videos for software training on editing tasks; 15 for image editing, 15 for text, and 15 for sound. This mix is to neutralise bias toward any kind of content that might favour video software tutorials with peculiarities. Fourth, I will determine the specific characteristics of each video using my framework. Fifth, I will run statistical analyses using SPSS to examine the relationship between video characteristics and video popularity. Finally, informed by literature and my findings, I will propose a set of guidelines for the design of instructional videos for software training published on YouTube that can hopefully aid practitioners when designing videos and researchers in prompting future studies into the effectiveness of video features for learning.

Theoretical Framework

In the following theoretical framework, I will define what YouTube's instructional videos for software training are. Further, I outline what relevant instructional design characteristics of these videos are, considering their peculiarities and that of YouTube as a platform. I will describe how the popularity of the instructional videos for software training is assessed using publicly available user statistics from YouTube.

Instructional Videos For Software Training

Instructional videos for software training (called video tutorials or tutorial hereafter), teach viewers knowledge and skills necessary to execute tasks using software. Software tutorials are typically screencasts and show step-by-step how to complete a task (Loyd & Robertson, 2012). The written or spoken narration that usually accompanies video tutorials provides the viewer with additional information on what is happening on the screen (van der Meij & van der Meij, 2013; Lloyd & Robertson, 2012). What clearly sets video tutorials apart from other types of instructional videos dealing with software is that the former's main focus is on modelling task execution. In contrast, providing information about tool options is characteristic for reference guides (van der Meij & van der Meij, 2013). Equally, recorded lectures that show a speaker giving an in-person audience instructions on how to use software fall outside the scope of this research; they are likely to include interactions between students and the speaker and thus possess different features than the video tutorials that are analysed as part of this research.

Learning how to execute tasks with software requires the acquisition of procedural knowledge. Procedural knowledge involves knowing-how to do something, such as performing a yoga move or using a technical device (Anderson, 1982; Schunk, 1996). In contrast, declarative knowledge can also be defined as knowing-what or knowing-that (Colman, 2008). It is described as knowledge of factual information, such as names, concepts and their relation (Anderson, 1982; Schunk, 1996). According to Hong, Pi & Yang (2018), the central learning process for developing procedural knowledge involves firstly knowing declarative knowledge associated with the procedure, and subsequently knowing how to carry out the procedure. Learning procedural knowledge is therefore suggested to be more difficult than declarative knowledge (Anderson, 1995). Instructional design principles, defined as rules for the design of instructional message that have been shown to support learning in empirical research, therefore appear to be particularly relevant when constructing videos that teach procedural knowledge, such as software tutorials.

Software instructions in a video format offer three main advantages compared to the traditional print manual as described by van der Meij and van der Meij (2014). First, video supports the presentation of multimedia, i.e.synchronous presentation of information in an auditory and visual mode. According to dual coding theory (Clark & Paivio, 1991; Paivio, 1986), the dualmodal presentation of information leaves stronger memory traces in the brain (Neelen & Kirschner, p. 233), which supports learning. Further, dual modality can partly overcome issues that occur when information is only presented in one mode (Mayer, 2014). Second, viewers can easily see and mimic the exact same actions that are performed in the video; video thus provides an “easy-to-follow model” (van der Meij and van der Meij, 2014, p.151). Third, video tutorials can be recorded as a screencast and thus display the identical environment and procedure that users have to engage in when completing the software tasks on their own; video also shows the dynamic movements that occur in the software and the passage of time (van der Meij and van der Meij, 2014). Consequently, video tutorials are congruent with the real-life task users have to engage in, which is found to enhance learning (Tversky, Bauer-Morrison & Betrancourt, 2002).

Nevertheless, research that compared video to written software instructions showed mixed outcomes. Some studies have findings that favouring paper instructions (Payne, Chesworth and Hill, 1992), while other studies have found no particular advantage for either of instruction type (Alexander, 2013) or they reported better training results for video but better test results for written instructions (Palmiter, Elkerton, Baggett, 1991; Palmiter and Elkerton, 1993). In contrast, Spannagel, Girwidz, Loethe, Zenlder and Schroeder (2008), Lloyd and Robertson (2012), van der Meij and van der Meij (2015) found better results for video. As pointed out by van der Meij and van der Meij (2015), a potential cause for the inconclusive results especially by Payne, Chesworth and Hill (1992) and Palmiter et al. (1991) could be that silent video was used in the experiments- an underuse of the multimedia potential of video. This also points to a general issue in video design and research, namely that video is often not designed in line with empirically proven design principles (Fiorella & Mayer, 2018); for this reason, Fiorella and Mayer (2018) suggest not to ask whether video is more effective than other instructioanl methods, but highlight the need for studies that identify which specific video features maximise learning with video.

Analysing The Characteristics Of Instructional Videos For Software Training

Videos possess a variety of characteristics based on which they can be analysed. The focus of this study is laid on characteristics that support learning, i.e. instructional design principles known of literature. Accordingly, a literature review was conducted to determine which instructional principles and frameworks are available and relevant for the design of instructional videos for software training.

Dual Coding Theory

The dual coding theory (DCT) by Paivio describes how the human brain processes verbal and visual information and postulates that the combination of words and images enhances learning (1986). DCT assumes that the human brain has two distinct yet connected cognitive processing channels, one for processing and representing verbal objects (i.e. written or spoken language) and a second one for nonverbal objects (e.g.. imagery and images). Information processing takes place in the corresponding channel, leading to the creation of separate representations and mental codes of the information. The created mental codes enable learners to remember and recognise the information. Most importantly, DCT (Paivio, 1986) proposes that the human brain codes written and spoken words once, but images are coded twice: first visually (e.g. a picture of a woman) and then verbally (e.g. the word ‘woman’). This means that when learners are presented with clearly relatable words and images, the “images leave double, integrated and thus stronger traces in our brains” (Neelen & Kirschner, 2020, p.233), which supports memory construction and enhances learning. As video can present both words and pictures, DCT gives us first insights into why video can be beneficial for learning.

Cognitive Load Theory

Another central theory for understanding how to design effective instructional material is Cognitive Load Theory (CLT). CLT is based on knowledge of the human cognition and the general learning process, and how manner in which information is presented can influence it (Sweller, 1988). CLT assumes that the human memory has three interrelated components: the sensory, working and long-term memory. Learning is successful when the information is encoded in the long-term memory, and the process is as follows: The sensory memory is transient and collects information from the environment which may then be temporarily stored and processed in the working memory, which has a very limited capacity to process information. Information processing in the working memory is a prerequisite for storing and encoding information into the long-term memory, which has nearly unlimited capacity. The central hypothesis of CLT is cognition is restrained by the limited working memory capacity (Sweller, 1994); in other words, the working memory capacity is the limiting factor in the learning process. CTL further suggests that information induces three different types of cognitive load, which partly can be managed to some extent.

Intrinsic load refers to the material-inherent load, which is determined by complexity of the material and the learners' prior knowledge (Sweller, 1994). In other words, intrinsic load is described as the "experienced difficulty of the subject matter" (de Jong, 2010, p. 107). A common example to illustrate intrinsic load stems from language learning: a simple translation of isolated words to another language (e.g. mouse= ratón, cat= gato) may induce a relatively low cognitive load, while translating complex, interrelated words such as sentences (e.g. the mouse ran from the cat) may induces a higher cognitive load- depending on the prior knowledge of the learner. For learners with little knowledge of the foreign language, translating the sentence may induce high load. In contrast, if learners are already familiar with the language, if they already know the words, how to conjugate verbs and so on, the task may induce lower cognitive load for them. This is the case because processing information that is already known, i.e. it is already stored in long-term memory, induces lower cognitive load than novel information.

Extraneous load is the second type of load and- unlike intrinsic load- it refers to the load caused by the manner in which information is presented and by the instructional procedure learners have to engage in (Sweller, Merriënboer & Paas 2019). Consequently, extraneous load can be managed by manipulating the used instructional material or procedure (de Jong, 2010). The third and final cognitive load type is germane load. Germane load was initially defined as the load induced by the learning process (de Jong, 2012), but it is currently assumed to redistributing working memory resources from processing extraneous to intrinsic information (Sweller, Merriënboer & Paas 2019).

Most importantly, CLT proposes that the total sum of all loads that information induces is decisive for learning: e.g. when the total cognitive load a video induces exceeds the learner-specific processing capacity, learning is impeded (Sweller, Merriënboer & Paas 2019). According to Sweller, Merriënboer and Paas, the two best strategies to regulate cognitive load are reducing extraneous load and optimising germane load (2019). More specific advice on how to employ these strategies in the design of instructional material like video is provided by the Cognitive Theory of Multimedia Learning and the Multimedia Principles (Mayer & Moreno, 2003).

Cognitive Theory of Multimedia Learning

The Cognitive Theory of Multimedia Learning (CTML) builds on the CLT and suggests how the brain processes multimedia messages (Mayer & Moreno, 2003). The term multimedia refers to combinations of words and pictures, such as a video with narration (Mayer & Moreno, 2003). CTML rests upon three assumptions: first, the human brain has two separate channels for processing information, an auditory and a visual one; second, each channel has limited capacity to process information; and third, learning is an active process that consists of "filtering, selecting, organising and integrating information" (Neelen & Kirschner, 2020, p.233).

In brief, the CTML's main premise is the multimedia effect, which postulates that under certain conditions, using multimedia material can enhance the brain's information processing capabilities. This effect can also apply to appropriately designed video. More specifically, a suitable combination of words and pictures enables the simultaneous processing of information in the auditory and the visual channel, which results in an overall higher information processing capacity. It supports the integration of information in existing cognitive structure (Brame, 2016) and cognitive load reduction, which can eventually lead to meaningful learning (Mayer & Moreno, 2003). Meaningful learning is desirable as it refers to high knowledge retention and the ability to transfer knowledge to solve problems in novel situations.

Furthermore, derived from empirical studies with multiple, randomised test, Mayer advanced 12 instructional principles for the design of effective multimedia learning material (2014). They are briefly presented in the following. Extraneous processing is similar to extraneous load in CTL (Sweller, Merriënboer & Paas 2019) and refers to cognitive load induced by processing words and pictures that are unnecessary for attaining the learning goal (i.e. extraneous material) or poor layout (e.g. embellishments, irrelevant details etc.). Given that working memory's capacity is heavily constrained, all processing capacity should be focused on processing essential material, i.e. words and pictures that need to be understood to attain the learning goal, rather than on processing extraneous material. Mayer proposes five Multimedia Principles that reduce extraneous processing (2014).

First, the coherence principle holds learning is enhanced when all unnecessary (extraneous) material is cut out so that the leaner's processing capacity can be used exclusively for dealing with essential material. Second, the signaling principle postulates that adding cues that highlight the organisation or structure of material facilitates learning as it helps the learner to focus processing of on essential, rather than extraneous, material. Third, the redundancy principle suggests to use pictures and narration, instead of pictures, narration and on-screen text; the latter would not aid learning and present redundant information that is better removed.

Two multimedia principles are suggested to prevent the split attention effect. When sources of information are split, the learner has to mentally integrate them to make sense of the material, a process that demands resources from the limited processing capacity of the working memory and leads to lower learning results. In contrast, sources of information should be presented in an integrated manner. The spatial contiguity principle holds that corresponding text and pictures should be physically integrated. For instance, labels should be placed close to corresponding picture elements rather than adjacent (Mayer, 2014). Similarly, the temporal contiguity principle postulates that learning is enhanced when narration and pictures are presented simultaneously, rather than in succession, as holding the information in the working memory would require unnecessary mental effort (Mayer, 2014).

Furthermore, three principles aid learners in managing essential processing so that an essential overload is avoided (Mayer, 2014). Essential overload takes place when essential cognitive processing needed for understanding the multimedia material exceed the learner's processing capacity. This can occur when material is complex and presented at a rapid pace. The segmenting principles holds that learning is improved when the material is presented in "learner-paced segments rather than as a continuous unit" (Mayer, 2014, p. 316). This principle is operationalised as chunking material and requiring users to press a continue button for proceeding to the next unit of information. It is assumed that the caused breaks are beneficial for learning because they allow users to fully process the information they were previously exposed to, before they chose to manually move on to the next segment and process the next segment of information.

Further, the pretraining principles is that learning is enhanced when names and characteristics of the central concepts of the multimedia lesson are known. This is based on the notion that once familiar with the main concepts and names, learners need to spent less processing capacity on them and can focus on other essential elements of the learning material. The modality principle indicates that learning is enhanced when information is presented in a dual mode fashion, that is visual information is paired with spoken, rather than written, words. Visual information is processed in the visual channel, while spoken words are processed in the auditory channel, making optimal use of the brain's processing capacities. Contrary, when visuals are combined with simultaneously presented written instead of spoken words, e.g. permanent on-screen text and no narration, all information is presented in only one, the visual mode. This would then require single channel processing, which leads to an overall lower cognitive processing capacity and lower learning.

The third and last category contains four multimedia principles that employ social cues. Social cues can increase the learners' perception of a personal relationship with the instructor, which enhances motivation and active cognitive processing and finally leads to an improved quality of learning outcomes (Mayer, 2014). The personalisation principle suggests that using a conversational rather than formal language style contributes to the learners feeling of having a relationship with the instructor. Employing a conversational style means that the instructor formulates message using "I" and addresses the learner with the frequent use of "you". The voice principle holds that learning is enhanced when the narration is spoken by a standard-accented human rather than a machine or accented human voice. Again, a narration spoken by a human contributes to the perceived social presence and eventually enhances learning outcomes. Moreover, the embodiment principle holds that learning is improved when “on-screen agents display humanlike gesturing, movement, eye contact and facial expressions” (Mayer, 2014, p.345). Contrary, the image principle postulates that seeing the narrator image that does not demonstrate much humanlike gesturing, moves or facial expression, may not necessarily contribute to learning.

Overall, the principles provide a foundation for understanding the cognitive processes involved in learning with multimedia. Yet, with regard to video design, the Multimedia principles offer rather general advice and can be described as “preliminary frameworks for building testable theories concerning instructional video” (Fiorella & Mayer, 2018, p. 469). For this reason, the literature research was continued to find more concrete design guidelines.

Eight Guidelines for the Design of Instructional Videos for Software Training

Van der Meij and van der Meij (2013) advanced ‘Eight Guidelines for the Design of Instructional Videos for Software Training’, hereinafter referred to 8G. The 8G are described as design patterns that offer ‘standard solution schemata’ (van der Meij & van der Meij, 2013, p. 206) for common issues in the design of software tutorials (Figure 1).

The 8G build on CLT and CTML and other theoretical or empirical evidence relevant for the design of video software tutorials. In short, the 8G provide a structured, condensed overview of strategies that support learners in acquiring and retaining software skills. In a comparative experiment, a video designed in line with the 8G led to more favourable outcomes than a paper tutorial on learner motivation, skill proficiency and skill retention (van der Meij & van der Meij, 2015). The guidelines and their subguidelines, as well as their main theoretical or empirical foundation, are briefly presented below.

Figure 1

Eight Guidelines for the Design of Instructional Videos for Software Training by van der Meij van der Meij (2013)

Guideline 1: Provide easy access 1.1: Craft the title carefully **Guideline 2: Use animation with narration** 2.1: Be faithful to the actual interface in the animation 2.2: Use a spoken human voice for the narration 2.3: Action and voice must be in synch **Guideline 3 Enable functional interactivity** 3.1: Pace the video carefully 3.2: Enable user control **Guideline 4: Preview the task** 4.1: Promote the goal 4.2: Use a conversational style to enhance perceptions of task relevance 4.3: Introduce new concepts by showing their use in context	**Guideline 5: Provide procedural rather than conceptual information** **Guideline 6: Make tasks clear and simple** 6.1: Follow the user's mental plan in describing an action sequence 6.2: Draw attention to the interconnection of user actions and systems reactions 6.3: Use highlighting to guide attention **Guideline 7: Keep videos short** **Guideline 8: Strengthen demonstration with practice**

Provide easy access. According to guideline one, users should easily find it easy to find information they are looking for. This notion is theoretically supported by Bethke et. al's (1981) criteria for user documentation. Van der Meij and van der Meij (2013) have identified that in the case of video software tutorials, the title plays a vital role in video search and selection process. An informative, easy to understand tutorial title makes it easier for the users to find it and check if the content meets their learning needs. The title should contain all relevant keywords and it should be phrased in such a way that it clearly outlines the tutorial goal, using a (gerund) verb and object structure based on Farkas recommendations for the logical and rhetorical construction of procedural discourse (1999).

Use animation with narration. Guideline two holds that a software tutorial should be designed as an animation with narration in line with Mayer's empirically proven multimedia principle (2005a), which holds learning with a combination of words and pictures leads to better learning than when words alone are used. Guideline two has three subguidelines that further specify how to optimise this combination of video and narration.

First, the video should show an authentic use scenario, which means that a software tutorial should be a screencast that shows how the task is executed in the actual software user interface. This is theoretically supported by Tversky, Bauer-Morrison and Betrancourt's (2002) congruence principle, which postulates that graphics are better understood and memorised if there is a coherence between the real-life object and the graphical representation of it. Additional empirical support for this subguidelines comes from Höffler and Leutner's (2007) meta-analysis on instructional animations, in which it was found that most realistic animation generates the highest learning outcomes.

Second, in line with Mayer's modality and voice principles (2014), van der Meij and van der Meij (2013) suggest using a human, standard accented voice for the narration. The narration should provide users with information on on what can be seen on screen and if it supports learning, other essential conceptual background information. Furthermore, in agreement with Mayer's redundancy principle (2014), the narration should not be repeated in written form as this presents redundant information and could overload the user's visual channel (Mayer, 2014).

The third subguideline holds that the video and narration should be synchronised in line with Mayer's temporal contiguity principle (2014).

Enable functional interactivity. Guideline three advocates to enable "functional interactivity" (van der Meij & van der Meij, 2013). Instructions should be designed in a manner that match the learners cognitive processing capacity (Kennedy, 2004; Mestre, 2012; Wouters, Tabbers, & Paas, 2007). Van der Meij and van der Meij suggest this match can be achieved by using two strategies: first, by optimising the video pace so that learners are neither bored nor overwhelmed by the speed in which information is presented. This notion is supported by studies from Bovair and Kieras (1991), Linek et al. (2012) and Mayer's segmenting (2014).

Second, offering user control functions is expected to enhance learning because it invites learners to actively engage with the video (Mayer, 2003). Further, this subguideline is supported by Tversky, Bauer-Morrison, and Betrancourt's (2002) apprehension principle, which holds that animations should easily understood. Because of the transient nature of video, this an be challenging as it puts high demands on the working memory; user control functions can be used to pause, stop and rewind video sections if needed (van der Meij & van der Meij, 2013). This is also in line with Mayer's segmenting principle which suggests that learning from video is enhanced when it can be adapted to the learner needs, rather than being presented as one continous unit (2014). Empirical findings from Schwan and Riempp (2004) also suggest that providing user control faciliates learning and reduces practices time for learning procedural knowledge and skills.

Preview the task. With guideline four, van der Meij and van der Meij (2013) propose to include a preview at the beginning of the video. A preview is a very brief, shortened version of the task demonstration in which the narrator shortly gives over the main steps of the task execution without elaboration. A preview before the actual task demonstration serves four goals: first, it prepares the viewer to be able to follow the demonstration as it helps them to build a general understanding and schema of the task execution; second, it draws attention to relevant aspect of the demonstration; third, it stimulates motivation and last, it reduces cogntive load during the actual demonstration. Guideline rests on Mayer's pretraining principle (2014).

According to van der Meij and van der Meij (2013), a good preview entails the three central aspects. First, the goal of the task is promoted; which theoretically supported by Farkas (1999) recommendations for procedural discourse. Second, the narration should be designed in a personal, conversational style using conversational language and frequently transmitting message in the first (I) or second person (you), which is in accordance with Mayer's (2014) voice principle and was also found to enhance learning in numerous other several other studies (e.g. Mayer, Fennell, Farmer & Campbell, 2004; Moreno & Mayer, 2000, 2004). Third, the preview should introduce critical vocabulary and concepts in the context of their use in line with the just-in-time principle proposed by van Merrienboer, Kirschner and Kester (2003). Furhter, introducing vocabulary and concepts in the moment of need faciliates learning because it reduces the working memory load during the actual task demonstration as the users will already be familiar with the concepts (van der Meij & van der Meij, 2013).

Provide procedural rather than conceptual information. Guideline five proposes to include more procedural-rather than conceptual- information in the tutorial. Van de Meij and van der Meij (2013) suggest that users turn to tutorials because they want to find out how to complete a software task. Therefore, the tutorial should provide the users with only the most essential procedural information necessary for immediate task achievement which is mainly procedural information. This notion is theoretically supported by the minimalism design principle, which advocates to support users in the completion of tasks (Carroll, 1990; Van der Meij & Caroll, 1998). Van der Meij & van der Meij further indicate that some conceptual background information may be needed the user to better understand what is happening, however, other superfluous information should be left out in order not to overload the users cognitive processing capabilities.

Make tasks clear and simple. Guideline six holds to present the demonstrated task(s) a in clear and simple fashion. In line with Mayer's coherence principle (2014), only the most essential information should be included in the video in order to avoid an overload of the user's working memory capacity. Therefore, the video should be kept as simple and short as possible by focusing on the demonstration of one relevant, practical task.

The tasks can be made clear and simple by focusing on three aspects. First, actions should be presented in the same natural sequence in which users have to engage in them. User actions can vary greatly depending on specific location, menu, tab, options or settings etc. that need to be used. This is referred to as condition or modifier of an action. As they orient the user, it is recommended to name them first and to then to go over to the action itself (Farkas, 1999), e.g. it should be stated on the "Layout Tab, click Breaks" instead of "Click on Breaks on the Layout Tab" .

The second subguidelines holds that the interdependence of the user's actions and the (software) system's reaction should be highlighted (van der Meij & van der Meij, 2013). Seeing the software's reaction is a form of immediate feedback, which has been found preferable in learning procedural skills, compared to delayed feedback (Shute, 2008).

Third, because a video tutorial presents the user with a stream of information to attend to, vierwers may have problems to see what is happening on the screen and thus fail to understand it. In accordance with Mayer's signaling principle, it is thus recommended to add visual clues that direct the user's attention to relevant information (2014). Further, pertinent information should be presented in a physically integrated manner in line with Mayer's spatial contiguity principle (2014). Further empirical support for cues is offered by research by De Konig, Tabbers, Rikers and Paas (2010), who found that cues in animations had a significant impact on user behaviour and learning.

Keep videos short. Guideline seven assumes that videos should be kept videos short in order to limit the amount of information users have to process and remember. In agreement with Mayer's segmenting principle, the user should be presented with a meaningful and manageable amount of information at once (2014). Van der Meij and van der Meij (2013) suggest that the video length is kept meaning full by showing one rounded off tasks with a clear beginning and end, i.e. the achievement of a goal or a sub-goal.

Strengthen demonstration with practice. Guideline eight postulates to offer practice files so that users can consolidate and improve their learning from video. Practicing their skills allows learners also to test whether they can apply their knowledge. Empirical findings by Ertelt (2007) indicates that learner performance was significantly enhanced after they were given the opportunity to practice their skills, compared to a no-practice control condition. It has to be noted though that learners in Ertelt's study did not have acccess to the instructional video during practice, a condition that van der Meij and van der Meij evaluate as not realistic in case of freely available videos (2013).

Overall, the 8G by van der Meij and van der Meij (2013) are a suitable basis for the video analysis in this research due to their specific focus on supporting the development and retention of software skills. However, it is important to note that the 8G were designed for a different use context. The original use context for the 8G was that a software manufacturer would use the guidelines to construct a set of tutorials and publish these on their website. The tutorials would guide the user from simple to more complex tasks and together the videos would form an instructional package (van der Meij & van der Meij, 2013). The present study focuses on another use context, namely that of tutorials published on YouTube. Consequently, an analysis of YouTube tutorials was undertaken to identify which elements of the 8G are relevant to assess the characteristics of software tutorials on YouTube, as well as how they can be translated into a meaningful framework that allows for the objective assessment of video characteristics.

Framework for analysis of YouTube's Instructional Videos for Software Training

In order to adapt the 8G framework to a YouTube context, several test analyses were run. Seven software tutorials were analysed in-depth and compared to the 8G in order to determine which elemenets need to be modified. A number of video features were mentioned by van der Meij and van der Meij (2013) in the description of the 8G, but not included as explicit sub-guidelines. Because these features have been found relevant in the video analysis, they were added as elements and are marked in light-blue in Figure 2. Moreover, the analyses demonstrated that a number of additional video features that relate to instructional design princples are present in popular software tutorials. Consequently, they were added as new elements and are marked in dark-blue, e.g. '2.3. Narrator Visibility'. For each of the sub-guideline, a number of manifestation were found. Each manifestation is described and illustrated in the results and discussion section and also shown in the codebook in Appendix 1. In the following, the rationale behind each of the newly added elements is outlined.

Figure 2

Framework for analysis of YouTube's Instructional Videos for Software Training, adapted from van der Meij and van der Meij (2013)

Guideline 1: Provide easy access
1.1: Title structure
1.2: Tutorial goal(s) and title match
1.3: Title specifies context
1.4: Title specifies condition

Guideline 2: Use animation with narration
2.1: Actual interface shown
2.2: Narration style
2.3: Narrator visibility

Guideline 3 Enable functional interactivity
3.1: Video pace
3.2: Pauses in narration

Guideline 4: Preview the task
4.0: Preview
4.1: Goal promotion
4.2: Conversational style in preview
4.3. Introduce new concepts
4.4: Goal illustration
4.5: Conversational style in goal illustration

Guideline 5: Provide procedural rather than conceptual information
5.0: Percentage of procedural content

Guideline 6: Make tasks clear and simple
6.1: Description of action sequences
6.2: Interconnections of user's action & system's reaction highlighted
6.3: Highlighting techniques

Guideline 7: Keep videos short
7.1: Video length
7.2: One rounded off task shown

Guideline 8: Music
8.0: Presence of background music

Van der Meij and van der Meij (2013) suggest to employ a (gerund) verb and object structure for the title (Farkas, 1999). The specific used title structure is assessed with item 1.1. Title structure. Further, the authors described that the title should be a matching description of all the goals that are achieved during the tutorial (van der Meij & van der Meij, 2013). This is evaluated with item 1.2. Tutorial goal(s) and title match.

In addition, further relevant aspects in the title design were identified bearing YouTube as the video source in mind. Guideline one to provide easy access is illustrated with examples from the website of software providers in the original 8G. Naturally, the user will expect tutorials from the providers on these websites. On YouTube however, anybody can upload a tutorial from any kind of software. Users seeking a tutorial about a specific software thus have to specify the software's name and possibly even the edition in the search query. Therefore, it can be beneficial to include the condition in the title and it is checked for with item 1.3. Tutorial specifies context. Furthermore, some software users seek a demonstration of how to execute a very specific task, under certain conditions such as the use of certain software functions or they want to obtain very specific results. This specification of conditions is assessed with item 1.4. Title specifies condition.

In contrast, the original guideline "2.3 Action and voice must be in synch" (van der Meij & van der Meij, 2013, p. 210) was removed as the pilot showed that this guideline was employed in all videos. One observation is that the narrator's image is frequently included in tutorials (Figure 3 and **Fehler! Verweisquelle konnte nicht gefunden werden.**). According to Mayer's image principles (2014), showing the instructor is a form of a social clue that does not always have a positive impact on learning. For the purpose of this study, it is determined if and in which parts of the tutorial the narrators are, which is assessed with item '2.3. Narrator visibility'.

Figure 3

Tutorial With Visible Narrator

Note. (source: https://www.youtube.com/watch?v=GbDnnnmH6UI).

Figure 4

Narrator Visible On Full Screen

*Note***.** (source: https://www.youtube.com/watch?v=PC4GRRCQPWE).

In line with the segmenting principle (Mayer, 2014), Van der Meij and van der Meij (2013) recommend to further to further aid the user in processing the presented information by including short breaks of 2-5 seconds after a subtask is completed. In these breaks, there should be no narration and no action performed on screen. Such pauses can give the user time to reflect on the shown subtask which increases learning from instructional video (Ertelt, 2007; Spanjers, Van Gog, & Van Merriënboer, 2012; Spanjers, Wouters, Van Gog, & Van Merriënboer, 2010). If and how breaks are included is evaluated with item '3.2. Pauses in narration'.

With guideline four, van der Meij and van der Meij (2013) suggest to include a video preview. While it was found a numbers of videos did not include a preview, tutorial goal promotions were present, using using different language styles. Given that goal promotions can enhance the learner's motivation and perception of task relevance, which can ultimately improve learning, they serve important instructional goals. Also, employing a conversational language style in the narration is important as demonstrated by Mayer's personalisation principle (2014). Therefore, the goal promotion and the employed language style were added as two new elements that are independent of a task preview, see items 4.4. goal illustration and 4.5. converstational style in goal illustration.

Guideline seven holds that videos should be kept short in order to limit the amount of information users have to process and remember. To do so, van der Meij and van der Meij (2013) recommend to limit the video length to the time it takes to demonstrate the achievement of a goal or a sub-goal. If this is done in the tutorials is evaluated with item '7.2. Show one rounded off task'. Furthermore, the total video length is measured in item '7.1. Video length'.

Van der Meij and van der Meij's original guideline eight holds to "strengthen demonstration with practice" (2013, p. 233) and postulates to offer practice files to users. This seems impractical on YouTube, as it is a platform made to share videos. In the video analyes, it was found that no reference was made to practice files, so the guidelines this feature was removed.

Instead, it was found that some tutorials contain music in various parts of the tutorial. While music in the introductory scenes may not be harmful for learning, music during the task demonstration could distract the users and therefore be extraneous material. According to Mayer's coherence princiiple, extraneous material should be weeded out as much as possible (2014). The presence of background music during task demonstration is checked for with item '8. Background music'.

Assessing video popularity on YouTube

YouTube automatically tracks and displays several publicly available user statistics that allow for the assessment of video popularity. Shoufan (2019) developed the 'Video Cognitive Value' (VCV) formula in order to assess the quality of YouTube videos designed for learning based on the number of views, likes and dislikes. The term "video cognitive value" refers to the "level of understanding provided by the video" (Shoufan, 2019, p.452). The formula was designed based on the results of a survey among 428 university students and it was examined which factors would lead viewers to like or dislike educational videos. About seventy-three percent of respondents stated they would like an educational video because they felt they understood the subject matter. In contrast, 47,6% of respondents indicated that they would dislike an educational video if they felt it did not help them to understand the topic (Shoufan, 2019). Other reasons for liking or disliking a video included mainly general video characteristics, such as video and audio quality, and video length, to name a few examples. Based on the study results, Shoufan developed the VCV formula displayed in Figure 5.

Figure 5

Formula for the assessment of video popularity by Shoufan (2019).

$$VCV = \frac{W_L \times N_L - W_D \times N_D}{N_V}$$

WL refers to cognitive weight of likes. To calculate this number, the number of likes (NL) is multiplied by 0,733 for the 73,3% of people liking a video because of its educational quality. WD refers to cognitive weight of dislikes. To calculate this number, the number of dislikes (ND) is multiplied by 0,476 to account for the 47,6% of people disliking a video because of its educational quality. NV describes the total number of views. Shoufan (2019) explicitly advises not to use this first formula because it does not account for the number of 'hidden dislikers'. Hidden dislikers are users that do not like the instructional video but do not press the dislike button, they simply skip the

Figure 6

Video popularity score Formula based on Shoufan (2019)

$$VPS = \frac{W_L \text{ x } N_L - W_D \text{ x } N_D}{N_V} \text{ x } (10^4)$$

Research Goals

This study seeks to identify which instructional design characteristics, hereafter referred to as 'characteristics', are present in YouTube's popular instructional videos for software training. Video popularity is determined using the VPS formula which uses input from publicly available video user statistics (views, likes and dislikes). The video characteristics are determined by coding the videos with a codebook that is based on the newly developed *'Framework for analysis for YouTube's Instructional Videos for Software Trainings*', hereinafter referred to as framework.

Finally, as a results of the video and statistical analyses, this study seeks to propose guidelines for the design of instructional video: one the one hand, a set of guidelines that support the theoretically effectiveness of the video as a learning tool; on the other, guidelines that, as suggested by the study's statistcal findings, support video popularity.

Consequently, the research questions are:

1. Which video characteristics, detailed in the framework, are present in the sampled videos?

2. How do the video characteristics determined with the framework relate to video popularity, as calculated using the VPS formula?

3. Based on the video analysis with the framework, which recommendations can be employed to support video popularity, as determined using the VPS formula?

4. Which video characteristics from the framework can be employed to support the theoretical effectiveness of instructional videos for software training on YouTube?

Method

The methods sections details about how this research was executed. First, the study design is illustrated, followed by an explanation of the sampling procedure and the general characteristics of the study subjects, i.e. the videos. Subsequently, the used measures and instruments are briefly described, i.e. Framework for Analysis for YouTube's Instructional Videos for Software and the VPS formula for gauging video popularity. Finally, the data analysis is portrayed.

Research design

The main research goal is to determine the specific characteristics present in the sampled instructional videos and how these impact video populartiy. The construct 'video characteristics' refer to a set of qualitative features, while the construct 'popularity' refers to a score that is based on quantitative, numeric variables (the number of views, likes, dislikes a video received). Given this combination of quantitative and qualitative constructs, this study employs a mixed methods research design (Johnson & Onwuegbuzie, 2004). The rationale for using both qualitative and quantitative data is that, if relationships between the constructs video popularity and individual video characteristics can be established, the results can inform video designers which video characteristics to employ in an attempt to influence their video's popularity on YouTube.

Sampling Procedure and General Subject Characteristics

A criteria-based sampling procedure was applied for two reasons. First, in order to find and select appropriate video software tutorials and not other type of instructional videos dealing with software. Second, further criteria were used to safeguard that only popular videos were sampled so that the VPS formula could be applied. The precise sampling procedure as well as the general subject characteristics are described below.

Step 1

To find the tutorials, the search bar on YouTube.com was used. The website was opened up using a browser in incognito mode to safeguard that search results were not personalised. The incognito mode allows for a neutral match between search terms and results because the findings are not affected by the past browsing history, networks or other previously saved settings.

In order to avoid a bias "toward one kind of content that might encourage instructional video with particular features" (Swarts, 2012, p. 197), a set of 15 software tutorials across three different task categories were gathered, namely image, sound and text editing. To design the specific search terms, guideline one from the 8G was applied; it postulates that the title should consist of a verb in gerund form and an object. Accordingly, the following search terms were framed: adding page numbers (to find text editing tutorials), removing background noise (to find sound editing tutorials), improving image resolution (to find image editing tutorials). These tasks are considered specific, yet basic software activities and they are deliberately free of the name of any kind of software product in order not to bias against tutorials product specific features. It was also expected that these search terms would yield a sufficient number of search results.

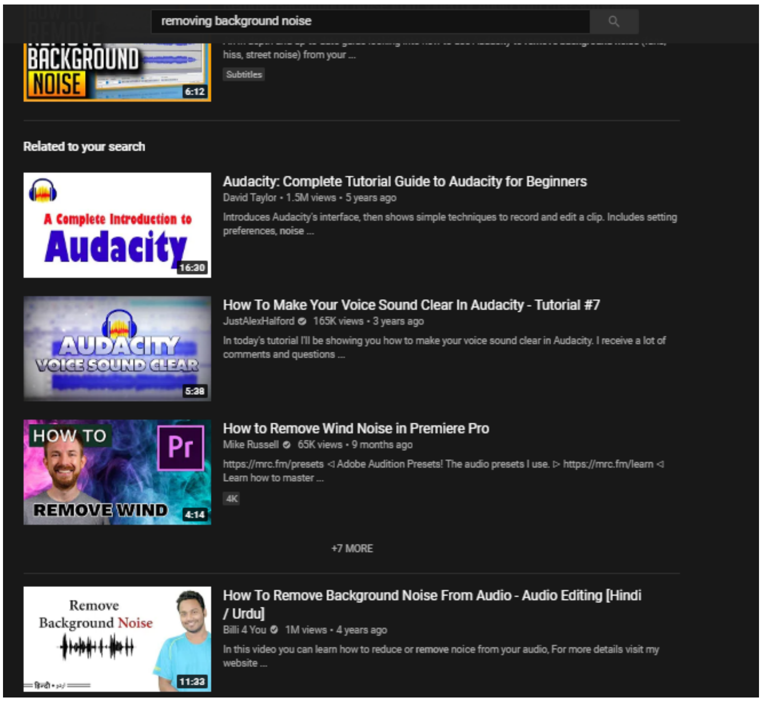
The rationale for using these terms in data sampling was that the tutorials show the task demonstration with a computer using a software in one of the three categories, text, image or sound editing. Accordingly, tutorials were selected even if they did not show the exact same task that was specified in the search terms, as long as they were still clearly demonstrating an editing task in said categories that was executed with software and a computer, not a smartphone.

Step 2

The search was started by typing in the search terms per task category. In order to obtain only the most popular videos per editing task, the filter option "view count" was selected. The search hits were then examined for their title and the number of views. In addition, a brief perusal of the video content was undertaken to ensure that it falls into the editing categories. The search hit list on YouTube has several features that cause the search hit list to be split up into different segments (Figure 7). It is unclear why these videos appear in the special segments. Videos from these segments were therefore not selected to be part of the sample; instead videos that appear above and below were checked.

Figure 7

Segmented search hits list on YouTube.com



In order to be selected, the videos had to meet certain baseline criteria. To ensure that the video popularity could gauged reliably, only videos with at least 1,000 views and a minimum total of 50 appraisals were considered. In addition, the videos had to be in English. To add, only one video per channel per task category was selected. If the search results showed several videos per channel, only the first video of that channel was chosen because videos from the same channel are likely to be designed in a similar fashion. Given the relatively small sample size, selecting several videos from the same channel would have increased their weight relative compared to all samples and therefore skewed the results.

After the search results were examined for all these criteria, the first 15 videos that met the criteria were selected. Subsequently, the videos descriptive data (e.g. URL, video title, video length, auto transcript etc.) and statistical data (number of views, likes, dislikes etc.) were stored in an excel file, as demonstrated in Appendix 1. Finally, the videos were downloaded using the software '3D YouTube Downloader', version 1.19.2.

Instruments

In order to analyse the sampled videos, several video characteristics were documented. Two instruments were used.

Descriptive and user statistic video data

In order to identify each video and to be able to measure video popularity with the VPS, descriptive and user statistic data of all videos was documented upon collection. This included a total of nine items: URL, video title, channel name, video length, number of views, likes and dislikes, upload and download date.

Framework for analysis of YouTube's Instructional Videos for Software Training

The framework for analysis was used to determine the characteristics of each sampled video. The framework consists of 22 items. Based on the framework, a codebook was created which describes each item in detail, illustrates indicators and provides examples so that videos can be objectively coded. The codebook can be found in Appendix 1.

Data analysis

In order to analyse video popularity, video characteristics, and the relationship between the two, a number of data analyses were performed.

Quantitative Data Analysis

To determine video popularity, a unique popularity score was calculated for each video, using the VPS fomula and the video -specific user statistic data (number of views, likes, dislikes; upload and download date).

Qualitative Data Analysis

To be able to assess the characteristics of each of the sampled videos, a video codebook was developed based on the framework. The coding categories of the codebook partly emerged from the literature review and partly from a pilot study. The codebook (Appendix 1) included nominal, ordinal and scale variables. The units of analysis were the the videos and the video transcript excerpts.

Interrater reliability. To ensure the reliability of the codebook, the codebook was tested using interrater reliability statistics. The testing process was the following. I trained a second coder, a junior educational scientist, by explaining the codebook items to him in a meeting. Subsequently, we independtly watched and coded the same set of four randomly selected videos from the sample. Then, results were compared and the interrater reliability for each codebook item was calculated using SPSS. An interrater reliability score of 0,61 was set as a minimum score a codebook item had to reach to be accepted. If an item did not reach this score, the codebook item was revised and the coding process was repeated until all codebook items reached the desired interrater reliability score.

The final results of the interrater reliability test are the following: For all nominal variables, a Kappa score of 0,98 was achieved. For all ordinal variables, Cohen's weighted Kappa κ= 0,84 was found, and for all scale variables a Pearson's correlation of r=0,92 was attained. According to McHugh (2012), an almost perfect agreement between coders was reached on the basis of the final codebook.

Mixed Methods Data Analysis

In order to determine the relationship between video popularity and video characteristics, the results of the quantitative and qualitative data analysis were used to perform a number of statistical test using SPSS. To start with, it was found that the data did not follow a normal distribution. Consequently, non-parametric statistical tests were run: Kruskal- Wallis and Mann-

Whitney tests where performed for categorical variables. For ratio variables, Kendall's tau-b correlations were computed. All analyses tested for a two-sided alpha set at 0.05.

Results and Discussion

In the following, the main results of quantitative and qualitative analyses per codebook item are presented along with a discussion of the findings. A full overview of the distribution of video characteristics per video category (i.e. image, sound and text editing) can be found in Appendix 1. Results and discussion are integrated to support the readers in comprehending the findings.

Raw Scores of the Variables in the Video Popularity Score Formula

The statistics of the variables included in the VPS formula are presented in Table 1. The data show that the sampled tutorials were watched by hundreds of thousands of viewers. Moreover, the data confirms several findings by Ten Hove and van der Meij (2015). To start with, only a small percentage of viewers left an appraisal in the form of a like or dislike. Second, likes (0,004%) were noticeably more common than dislikes (0,0003%), similar to Ten Hove and van der Meij´s findings (2015). Third, the standard deviation for views, likes and dislikes was higher than mean scores for these variables. As pointed out by Ten Hove and van der Meij (2015), this indicates very large differences in range and frequency within and across groups. An example for the huge differences in range is that the least viewed tutorial was watched 62.380 times, compared to the most watched tutorial that has 22.522.657 views.

Table 1

Mean frequency (Standard Deviation) for Views, Likes and Dislikes.

	Sample including extreme outliers (*N*=45)	Sample excluding extreme outliers (*N*=43)
Views	1.017.680 (3.302.024)	1 017 679,96 (3 302 024,182)
Likes	3.933 (5.265)	3 932,84 (5 265,252)
Dislikes	329 (335)	329,36 (335,912)

Table 2

Descriptive Statistics of video tutorials split per task category

Type of software task		N	Minimum	Maximum	Mean	Std. Deviation
Image editing	Video popularity score	15	8,25	275,18	49,88	68,68
	Views	15	62380	913686	342207,47	272365,85
	Likes	15	148	32000	3481,13	7968,06
	Dislikes	15	24	1100	307,53	352,52
	Days online	15	492,00	3741,00	1697,13	944,67
Sound editing	Video popularity score	15	8,23	198,37	68,91	48,89
	Views	15	274648	775491	403352,33	142002,60
	Likes	15	1200	8700	3813,33	2743,52
	Dislikes	15	57	507	206,33	117,17
	Days online	15	421,00	3767,00	2102,1333	1001,85
Text editing	Video popularity score	15	4,38	82,62	27,90	19,07
	Views	15	396177	22 522 657	2307480,07	5612285,27
	Likes	15	761	14000	4504,07	3940,63
	Dislikes	15	72	1500	474,20	421,22
	Days online	15	927,00	4163,00	2301,60	948,61

The comparison of the descriptive statistics of the tutorial when split per task category shows that text editing tutorials have the lowest mean popularity score, is on average online for the longest time period, and has the most views. Sound editing tutorials have the highest mean popularity score and the second highest view count. Given the small sample size,

Guideline 1- Enable easy access

1.1. Title structure

Results. The analysis of the title structures showed the vast majority of tutorials (62,2%) followed a 'how-to title' structure'. An exemplary tutorial title is "How to insert different page numbers in Word". Further, 13,3% of videos complied with the 8G and used a 'verb and object' title structure. An example of this recommended title structure from the sample is "Improve low resolution images quality in Photoshop cc 2017". To add, none of the sampled videos adhered to the second recommend title structure 'gerund-verb and object'. Finally, the remaining 24,4% of videos had other title structures, e.g. "Low to High Quality/Resolution Photo/Image in adobe Photoshop", "Noise Reduction with Edison". A Kruskal-Wallis test revealed that there was no significant difference in video popularity across title categories, H (2)=0,388, p=.824.

Discussion. The title structure analysis of YouTube's popular instructional videos is to the best of the authors knowledge the first of its kind. Previous studies by Swarts (2012), and Swarts and Morrain (2012) did not inspect this characteristic.

This study's results showed that the in literature proposed title structure of verb plus object, which is recommended because it allows for a concise description of the tutorial goal (van der Meij & van der Meij, 2013), is not commonly used in the sampled videos. In addition, the title structure of gerund verb plus object, described as the optimal choice in procedural discourse because the gerund verb transmits the active, procedural nature of the task (Farkas, 1999), was not present in the sample. In contrast, tutorials mainly had a how-to title structure, which is unsurprising, given that videos for procedural knowledge development are often nicknamed how-to videos. Although the how-to title structure deviates from the theoretical recommendations for titles from the streamlined-step model (Farkas, 1999), the how-to title still serves the title's most important goal to transparently convey the purpose of the tutorial. Further, it was striking that the second most common choice for tutorial titles was to use the 'other' title structure; this included so-called noun titles, such as "Word 2013: Headers, Footers, and Page Numbers". Farkas (1999) describes this title type as the least informative title structure: in the title "Word 2013: Headers, Footers, and Page Numbers", for instance, it is not clear what the tutorial demonstrates just by reading the title: Does it give an introduction into what headers, footer and page numbers are? Or does it show how to add or adapt them? This remains unclear because a verb that describes the actual tasks is missing in this type of title.

Despite the fact that the title structure was not found to relate to video popularity, an informative title structure in the how to or (gerund) verb plus object format is important becaues it helps the users understand which task the video is showing. Knowing this, users can make an informed choice about whether the video meets their learning goal or if they have continue their search.

1.2.Title match and tutorial goal(s) match

Results. The title should be a concise description of all the goal(s) the tutorial demonstrates. To assess if this was the case, the title was compared to all the tasks that were performed in the the sampled videos. If there was a deviation, e.g. more or less tasks than indicated in the title were shown, the title was assessed as showing an unclear or no match between tutorial goals and title. Examples of such deviations are shown in Table 3. Most of the tutorials (77,8%) had a clear match between the tutorial goal(s) and the title. An unclear or not match between tutorial goal(s) and the title was found in 22,2% of tutorials. A Mann-Whitney-U test showed that there was no significant relationship between tutorial goal(s) and title match, and video popularity, $U(N_{\text{tutorial goal(s) and title match}}=35, N_{\text{unclear or not match between tutorial goal(s) and title}}=10) = 120$, $p=.133$.

Discussion. It is noteworthy that image and text editing tutorial had a larger percentage of tutorials with a clear match than sound editing tutorials; i.e. 93,3% and 80% of match for image and sound editing tutorials respectively, compared to a match of 60% for sound editing tutorials. The lower percentage of match between tutorial goals and titles of sound editing tutorials was caused due to the following issues: without specifying it in the title, a number of sound editing tutorials showed numerous noise removal techniques, demonstrated the removal of different noise types, or showed the task completion using more software programs than indicated. The coverage of tutorial goals in the title was therefore not given.

Van der Meij and van der Meij further highlight the importance of a title that fully covers all shown goals because a matching title is more likely to evoke appropriate user expectations about the video content. The implication for designers is to choose a title that represents “the most general action” (Farkas, 1999, p. 46) and yet makes the goal(s) clear. In case it is not possible to find a concise and matching title because several tasks are performed, the tutorial might be best split into several individual videos.

Table 3

Examples of Unclear or No Match Between Title And Tutorial Goal(s)

Tutorial title	Design issues
Insert Sections with Different Page Numbers (Word)	Tutorial demonstrates how to insert and format page numbers, but this goal is not indicated in the title.
Premiere Pro TUTORIALS - How to Remove Audio Noise/ Background Noise	Tutorial demonstrates a second technique for removing background noise using another software than Premiere Pro, which is not indicated in the title.
How to Increase Resolution - Affinity Photo Tutorial	In addition to showing how to increase the resolution of a phot demonstrates how to sharpen it without indicating this in the ti

Word 2013: Headers, Footers, and Page Numbers	The title does not include action verbs and is vague, it is unclear what it shows. The tutorial actually only shows how to adapt the header but not the footer.

1.3. Title specifies context

Results. Title specifies context refers to whether a tutorial title included the software name and edition. An example of a title that specifies the context by stating the software name and edition from the sample is “Adobe Audition CS6- How to Remove Noise from a Clip”. More than half of the tutorial titles (57,8%) included the software name and more than a quarter of tutorials titles (26,7%) included both the software name and edition. Further, 15,6% of tutorials did neither include the software name nor edition. A Kruskal-Wallis test revealed that there was no significant difference in video popularity and categories of tutorials specifying the context, H (2) =3,628, *p*=.163.

Discussion. Including important keywords in the title can facilitate user access to the tutorial (van der Meij & van der Meij, 2013). An example of important keywords can be the software name and edition because they specify the use context of the tutorial, i.e. the task execution using a particular software. The majority of tutorials included either the software name or the software name and edition.

While including the software name and title did not influence video popularity, including them offers two important advantages with regards to usability and easy access of to the video. First, because interfaces and procedure differ among software programs, tutorials are likely to be useful only to users of the same software. Clearly signaling in the title which software is used could attract the relevant user target group that utilises this software, and it signals to user of other software that they may need to look for another tutorial. Second, including the software name in the title has an advantage over writing it in the video description box which is shown below the video. The video title is displayed in a larger font size in the search hits list, whereas video description is shown in a smaller font. Including the software name and potentially also the edition in the title thus means that this key information is presented prominently, so it may be a more perceptible to users.

1.4. Title specifies condition

Results. Tutorial titles can also contain conditions, that is very specific end results that are achieved by employing the demonstrated procedure. For instance, a text editing tutorial may show how to start page numbers at a specific page or how to use a specific software functions such as page breaks. The analysis showed that the 64,4% of tutorials did not include such a condition, while 35,6% did specify a condition in the title.

Discussion. This finding is likely to be influenced by the search terms that were used in the video sampling procedure; the YouTube search mechanism matches search terms with video titles. General search terms without specifications were used and because of this potential bias towards tutorials without specific conditions, it was not calculated if this video characteristics relates to video popularity.

As indicated by Farkas (1999, p.46), the title should represent "the goal state that the procedure will result in" but at the same time also "the most general action ", that is shown in the tutorial. Video designers should consider including the condition in the title if their tutorial mainly revolves around a condition, or when a condition is an important keyword (van der Meij & van der Meij 2013; Farkas, 1999). The potential disadvantage of including the condition is that it can result in long video titles. Literature does not provide specific recommendations about the maximum title length; video designers are advised to carefully weigh the benefits of specifically stating the condition in the title against choosing a more general, shorter title.

Guideline 2- Use animation with narration

2.1. Show the actual user interface

Results. Most tutorials show the actual user interface as indicated in the tutorial title or narration (91,1%). In 6,7% of the tutorials, it was unclear if the actual user interface was shown because the used software was neither clearly mentioned in the title nor in the narration. Only one video (2,2%) showed a user interface other than indicated in the title

Figure 8. This tutorial showed the task execution in an older software edition (CorelDRAWX4), although the tutorial title clearly specified a more recent edition (CorelDRAWX5). A Kruskal-Wallis test revealed that there was no significant difference in video popularity and showing the real user interface, H(1)=3,665, p=.056.

Figure 8

Tutorial with Incongruent User Interface

-*Note*. (source: https://www.youtube.com/watch?v=10h354ivjW8).

Discussion. Showing the actual interface in the tutorial is beneficial for learning because a resemblance of the graphical representation and the actual object helps learners to understand and remember the graphical representation (Tversky, Bauer-Morrison & Betrancourt, 2002). In the case of software tutorials, the actual object refers to the interface users are going to see when attempting to complete the task on their own; the more realistically the tutorial shows this interface, the better users can perceive and remember it. Not clearly stating which interface is shown, or showing a different interface, is unfavourable for several other reasons. Figure 8 shows a tutorial with a discrepancy between stated and used interfaces. After seeing that the tutorial shows the task demonstration in edition X4 and not X5 as claimed in the tutorial title and onscreen text, users may stop watching and look for another video instead. At least, they are likely to be puzzled by this discrepancy. Moreover, there can be differences in the user interface between software editions, so instructors may convey outdated or incorrect action steps when they do not use the exact same software edition they claim to use. In sum, although not related to video popularity, video designers are advised to show the actual user interface.

2.2.Use a spoken human voice for narration

Results. The analysis of the applied narration style demonstrated that 62,2% of the tutorials employed exclusively a human voice for narration. More precisely, 53,3% of the tutorials had a standard-accented narration and 8,9% had a non-native, foreign-accented narrator. A mixed narration style was applied in 17,8% of the tutorials, meaning they included both spoken and on-screen written narration. Further, 15,6% of tutorials included a narration only in the form of written on-screen text. Another 4,4% of tutorial did not include any narration but only showed the task demonstration. None of the tutorial employed a computer-generated voice for the narration.

To determine the relation between video popularity and narration style, a number of statistical analyses was run. A Mann-Whitney U test indicated that there was no difference between tutorials that had a narrator (i.e. spoken narration or mixed-style narration), compared to tutorials that did not have a narrator (i.e. no narration, on-screen text only), $U(N_{arrator\ present}=36, N_{arrator\ absent}=9)=93,00$, $z=-1,958$, $p=0.05$. To add, a Kruskal-Wallis test revealed that there was no significant difference in video popularity between tutorials employing the recommended human voice narration style and other narration styles, H (4) =5,958 $p=.202$.

Discussion. In order to and to make full use of the potential to supply synchronous presentation of information via the visual and auditory channel and thereby cognitive processing improve learning, it is recommended to employ a spoken narration using a human, standard accented voice (van der Meij & van der Meij, 2013). Further, Mayer's modality advocates the use of visuals paired with spoken words rather than on-screen text (2014). In certain instances, however, showing on-screen text can support learning, e.g. when the learner is presented with technical terms (Clark & Mayer, 2011). A screenshot of a sound editing tutorial which showed technical term is presented in Figure 9.

Figure 9

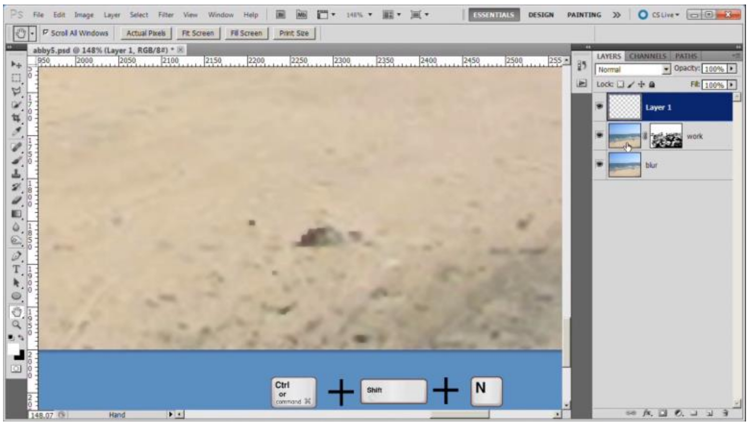
Tutorial with Technical Terms as On-Screen Text

Note. (source: https://www.youtube.com/watch?v=Z1CyFNoAWZc).

Another instance of technical terms in software tutorials are keyboard short-cuts, these were displayed as on-screen text in three of eight tutorials that employed a mixed narration style as depicted in Figure 10. The keyboard short cuts are displayed graphically similarly to their actual appearance on the keyboard. This congruence of the learning object and its graphical representation is generally conducive for learning (Tversky, Bauer-Morrison & Betrancourt, 2002; Höffler and Leutner, 2007), so showing the shortcuts in this way could theoretically enhance learning.

Figure 10

Tutorial with Mixed Style Narration With Keyboard Short Cuts as On-Screen Text



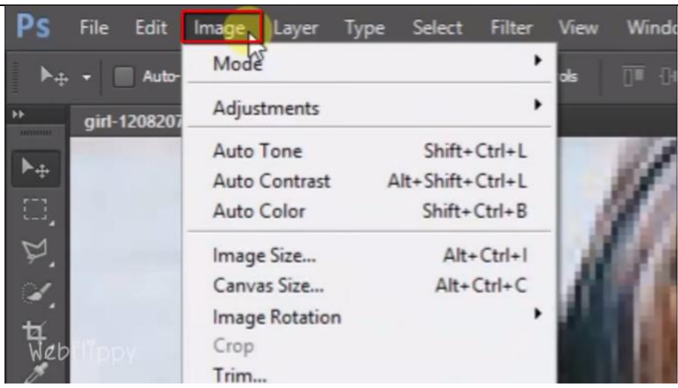
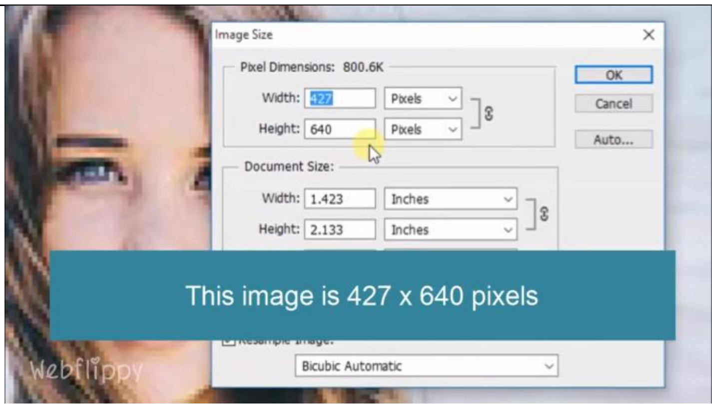
Note. (source: https://www.youtube.com/watch?v=CzFDKV9FDJg).

Strikingly, in 15,6% of tutorials there was no spoken narration and instead information was transmitted as on-screen text. Generally, this on-screen text was very sparse, ranging from a minimum of 6 to a maximum of 97 words per complete tutorial, with a mean of 38 words per tutorial. Table 4 shows an image-editing tutorial with such a very sparse written narration of six words in total. Given that no additional information provided by a written or spoken narration, it was perceived as difficult to follow and understand the tutorial. Often, in tutorials with such sparse narration, visual cues were used to convey information about action steps as shown in segment one in Table 4; signaling of the mouse cursor and frames were applied to convey the choice of menu options.

Using on-screen text instead of narration means that users have to process all information with their visual channel, which bears the risks of essential overload and consequently can have detrimental effects on learning (Mayer, 2014). Additionally, applying on-screen text negatively affects other instructional design principles such as the recommendation to highlight the interconnection of the user´s action and the system's reaction (van der Meij & van der Meij, 2013). Having to focus one's visual attention on the written text means that the viewer cannot observe other on-screen happenings that occur at the same time. The viewer may thus miss to observe the interplay of the user's action (e.g. selecting an option from the user interface) and the system's automatic response (e.g. opening up of additional user menu).

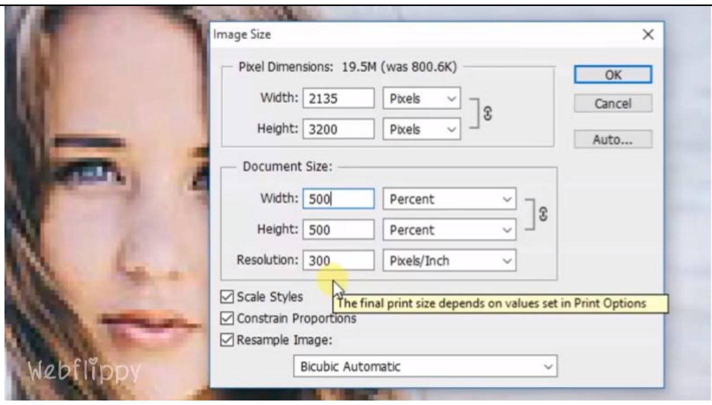
Table 4

Image Editing Tutorial With Very Sparse On-Screen Text As Narration

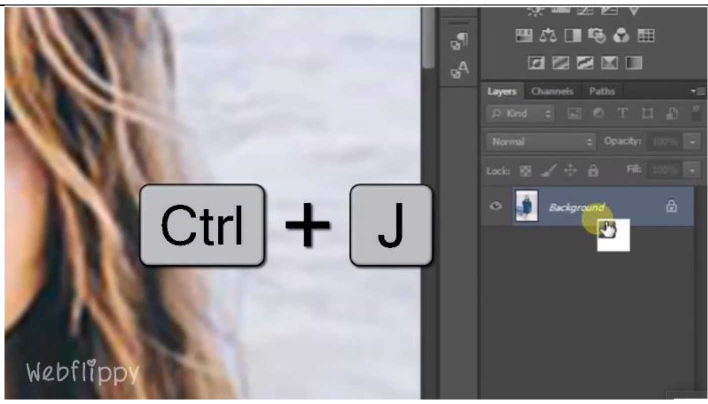
1. Photoshop Tutorial | How to Improve Low Resolution Photos in Photoshop No spoken narration Timing: 0.31	
2. Photoshop Tutorial | How to Improve Low Resolution Photos in Photoshop No spoken narration Timing: 0.36-0.39	

<table>
<tr>
<td>3. Photoshop Tutorial | How to Improve Low Resolution Photos in Photoshop

No spoken narration

Timing: 0.48</td>
<td>
</td>
</tr>
<tr>
<td>4. Photoshop Tutorial | How to Improve Low Resolution Photos in Photoshop

No spoken narration

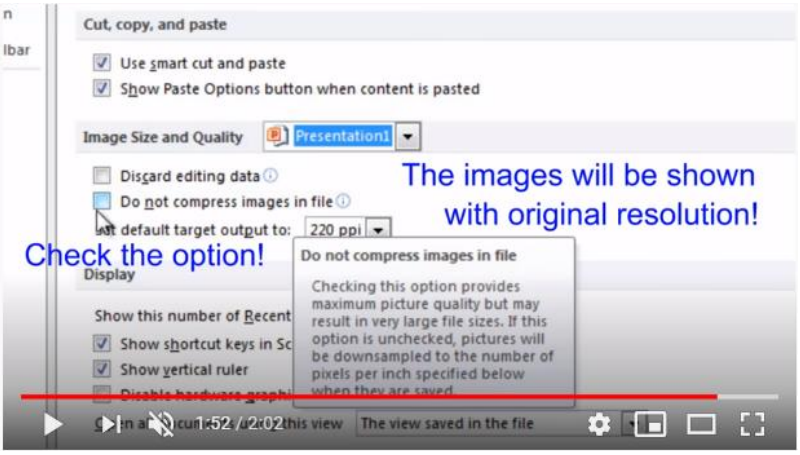
Timing: 01.10-01.11</td>
<td>
</td>
</tr>
</table>

Note. (source: https://www.youtube.com/watch?v=glktfwNDX_s).

Another example of example a tutorial using on-screen text instead of a spoken narration is shown in Figure 11. The shown written narration is not only problematic because it is placed on top of the user interface, thus interfering with a clear view on the task demonstration. It is also problematic because written text in videos can cause a type II essential overload (Mayer, 2014). This type of overload is caused when the learner's channel processing capacities are exceeded with essential information, in this particular video this is likely to happen because learners have to simultaneously follow the happenings on screen and read concurrent on-screen text, which could overload their visual processing channels.

Figure 11

Tutorial with Poorly Placed On-Screen Narration



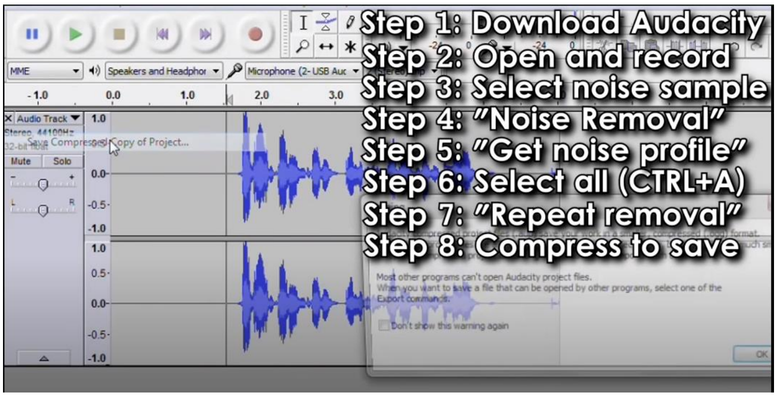
Note. (source: https://www.youtube.com/watch?v=NFGx78yFvy8).

In the mixed-narration style tutorial displayed in

Figure 12, action steps are shown as permanent on-screen text. While making the structure and organisation of multimedia material clear by highlighting it is beneficial for learning according to the signaling principle (Mayer, 2014), the struture presented in Figure 12 is unfavourable for learning because the text is placed so prominently on top of the user-interface that it is difficult to see the actual task demonstration. This demonstrates the importance of finding an adequate location for on-screen text. The text should not interfere with an optimal view on the task-demonstration.

Figure 12-

Tutorial poorly placed On-Screen Action Steps



Note. (source: https://www.youtube.com/watch?v=tAJ4Sg-nO6A).

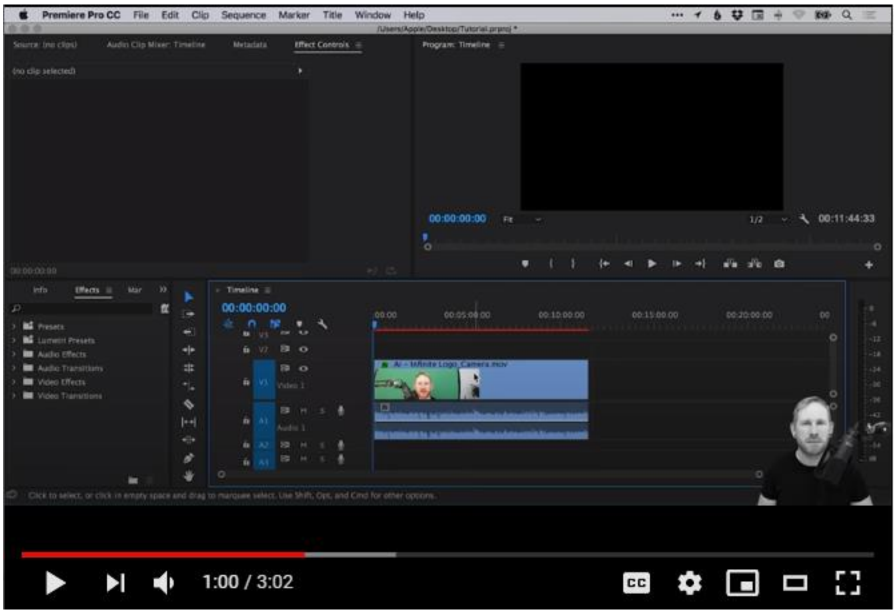
In sum, although not related to video popularity, a spoken narration is highly recommended as it supports learning and avoids many design issues related to on-screen text or a lack of narration. Video designers should be aware of the illustrated problems that can arise when a spoken narration is missing or on-screen text is used exclusively: it may cause cognitive overload and underuses the possibility of video. Designers may choose to display keyword or relevant keyboard-short cuts as on-screen text when this does hinder a clear view on the task demonstration.

2.3. Narrator visibility

Results. The majority of tutorials (57,8%) employed a voice-over, meaning that the narrator was not visible and only audible. In 13,3% of the tutorials, the narrator was constantly visible in a picture-in-picture montage as shown in Figure 13. In 8,9% of the tutorials, the narrator was visible only in task-demonstration free scenes, such as the beginning or ending scene, or when he or she was explaining conceptual knowledge to the viewer. Narrator visibility was not applicable to the remaining 20% of tutorials due to the fact these tutorials did not include a spoken narration.

Figure 13

Narrator Visible in A Picture-In-Picture Montage



-*Note*.(source: https://www.youtube.com/watch?v=hTIFREnZi34).

Several tests were run to assess how narrator visibility affected video popularity. First, a Mann-Whitney U test showed that tutorials showing the narrator were significantly more popular than tutorials that did not show the narrator, $U(N_{\text{Narrator visible in scenes without performed action steps and narrator permanently visible}} = 10, N_{\text{narrator not visible}}=26)= 30{,}000$, z=3,532, p=0.000. Post-hoc Mann-Whitney U tests with an alpha level of .05 were used to compare all pairs of groups and revealed two significant findings. To begin with, tutorials which show the narrator in scenes in which no actions steps are performed were significantly more popular than tutorials that did not show the narrator, $U(N\ N_{Narrator\ visible\ in\ scenes\ without\ performed\ action\ steps}= 4, N_{narrator\ not\ visible}=26)= 1{,}0$, z=-3,111, p=0.000. However, it is important to note that the number of cases in the category ´narrator visible in scenes without performed actions steps´ was smaller than five, so the experiment should best be repeated with a larger sample size to verify this finding.

Furthermore, tutorials which permanently showed the narrator were also significantly more popular than tutorials which did not show the narrator and had a voice over only, $U\ (N_{Narrator\ permanently\ visible}= 6, N_{narrator\ not\ visible}=26)= 29{,}000$, z=-2,366, p=.018. A significant difference in video popularity between tutorials that show the narrator in scenes in which no actions steps are performed and tutorials that permanently show the narrator was not found, $U\ (N_{Narrator\ permanently\ visible}= 6, N_{Narrator\ visible\ in\ scenes\ without\ performed\ action\ steps}= 4)= 9{,}000$, z=-0,640, p=0,522.

Discussion. A potential explanation for the increased video popularity in relation to a visible narrator comes from a recent study by van Wemeskerken, Ravensbergen and van Gog (2018) who examined the role of instructor presence on learning from video that models examples on how to resolve probability problems. In the experiment, college students were split into two groups; one was shown a video in which the narrator was constantly visible and the other group was shown a video without a visible narrator. Eye-tracking data indicated that students looked at the instructor for about 30% of and that their visual attention allocation towards the instructor did not change throughout the video. This finding indicates that human faces are particularly attractive social clues (Wemeskerken, Ravensbergen & van Gog, 2018), which could explain the viewers attraction and an increased video popularity of tutorials with a visible narrator.

With regards to the effect of showing the narrator on-screen, it is important to note that its effect on learning is not entirely clear. The study by van Wemeskerken, Ravensbergen and van Gog (2018) did not find any positive effects of instructor presence on learning outcomes. On account of the facts that learning did not improve with a permanently visible instructor and that it may distract learners from the learning material itself, the authors overall advice against taking the risk to showing the narrator's face instructional videos. Fiorella and Mayer (2018) further hypothesize that seeing the instructor's face may have positive effects for easy tasks; no effect for moderate tasks, and could have detrimental effects for complex tasks as the face takes away visual attention from other elements of the video, but systematic research is needed to assert these hypotheses.

In contrast, findings from Guo et al. (2014) suggest that showing the narrator in opportune scenes can have positive effects on learning. In their analysis of 6.9 million view sessions of MOOC videos, they found that displaying the instructor's head in suitable moments of the video lead to an improved engagement with the video, i.e. longer video watch time and more problem-solving attempts in a follow-up assessment (Guo et al., 2014). In software tutorials, opportune moments are the scenes in which no actions steps are performed, given that seeing the narrator's face could distract the users' visual attention away from performed actions. As displayed in Table 5, in the sampled videos scenes in which no action steps were performed included introduction or ending scenes, and scenes in which conceptual information was explained.

Overall, video designers interested in designing both effective and popular software tutorials are advised to show the narrator's face in task-demonstration-free scenes in order to avoid any risk in distracting the users' attention from essential the on-screen actions.

Table 5-

Narrator Visibility in Opportune Moments Of The Video Tutorial

1. How to Increase Resolution - Affinity Photo Tutorial Narration: Hello and welcome to an Affinity Revolution tutorial my name is Ezra Anderson and today we're going to increase the resolution of a picture. Timing: 0.00-0.10	
2. How to Increase Resolution - Affinity Photo Tutorial Narrator: We'll start off with a pixelated photo and then give it a higher pixel density. Timing: 0.11-0.14	
3. How to Increase Resolution - Affinity Photo Tutorial Narrator: Of course, you'd always be better off taking a high-quality picture right from the camera but if all you have to work with is a low-quality photo then this technique cannot least help make the picture a little bit better […].	

Timing: 0.14-0.27	

Note. (source: https://www.youtube.com/watch?v=PC4GRRCQPWE).

Guideline 3- Enable functional interactivity

3.1. Video pace.

Results. The basis for the assessment of the video pace was the narrative speed that was operationalised as words per minute (wpm). For the word count, the video transcript provided by YouTube was used and divided by the total tutorial length (Table 6). A mean score of 162 words per minute with a standard deviation of 31 was found for all videos with a spoken narration (*N*=36, range 93-224 wpm, *SD*=31). A Kendall's tau-b correlation was run to determine the relation between narrative speed and video popularity but no significant relationship was found, τ_b (*N*=36) = 0,102, *p*=0,383.

Table 6-

Video Paces Found In Dataset In Words Per Minute (wpm)

Distribution a Distribution across the whole sample		Distribution across tutorial types					
	Total (*N*=36)	Image (*N*=7)		Sound (*N*=15)		Text (*N*=14)	
Mean	162	Mean	145	Mean	170	Mean	162
Standard Deviation (*SD*)	31	SD	42	SD	24	SD	30
Range	131	Range	101	Range		Range	112
Minimum	93	Minimum	93	Minimum	113	Minimum	108
Maximum	224	Maximum	194	Maximum	224	Maximum	220

Discussion. In literature, it is recommended to use a conversational speed for the narration (Swarts, 2012; van der Meij and van der Meij, 2013). However, a concrete and widely agreed upon wpm range for conversational speed of native English speaker is not established. Despite, Tauroza and Allison (1990) found that the average speaking rate of adults in English ranges between 150 to 190 words, while Walker (2010) and Laver (1994) stress that this rate may rise to 200 wpm in conversations. For the purpose of this study, the conversational speed is operationalised as a range from 150 to 200 wpm. This applies to about 56% of tutorials with a spoken narration (*N*=36). Only 28% had a wpm rating below 150 wpm, and 17% a narrative of more than 200 wpm. The average narrative speed in the present study was slighter higher than that of popular instructional videos for declarative knowledge development (Mean= 145 wpm score, *SD*= 40) found by Ten Hove of van der Meij (2015).

Recent studies give more precise information about which kind of narrative speed to apply to support learning with instructional videos. Guo et al. (2014) found that a wpm range of 185-254 wpm generated highest engagement scores across learners in MOOC videos. Moreover, Miao, Zhang and Chen (2020) found that the learners' age is a critical factor to take into account when designing the speaking rate of instructional videos in a study that tested the effect of learner age, speaking rate and learner academic background in pre-posttest score. More specifically, Miao et al. (2020) found that the learners' age is negatively correlated with a preference for high speaking rate (210-240 wpm), irrespective of the learners' academic background. This means the older the learner, the slower the narration should be to support learning (Miao et. al, 2020). Further, Miao et. al (2020) found that learners without background knowledge about the topic of the video were least sensitive to changes in the speaking rate. For this type of learner, medium (180- 209 wpm) and high (210-240 wpm) speaking rates were similarly effective, which means that a broad range for the narrative speed for this type of learners is may be most suitable.

In short, in order to design an effective narrative speed, designers can either decide on which age group they would like to target with their tutorials and design the speed accordingly, or they can check the average age of the members of their YouTube channel. This is a function available to channel owners if channel members opted to publish their age on YouTube. Subsequently, video designers should consider the difficulty level of their video to see what learner type it attracts and accordingly apply a narrative speed in line with the findings of Miao et al. (2020), to optimise learning from their video.

3.2. Inclusion of pauses.

Results. Another feature pertaining to the video pace are pauses. As part of this study, pauses were operationalised as segments in which there is no narration in combination with a still screen, for a minimum duration of two seconds. This meant that only videos including a narration were evaluated with regards to this feature (*N*=36). The inclusion of pauses was then assessed on a seven-point Likert scale, ranging from 'never' to 'frequently'. None of the tutorials included pauses.

Discussion. The lack of pauses in tutorials disregarded the recommendations made in the 8G (van der Meij & van der Meij, 2013) according to which including pauses at suitable moments is beneficial for learning for two main reasons. First, they give learners time to process new information and to reflect on it, thereby controlling cognitive load (van der Meij & van der Meij, 2013). Second, pauses are described as important boundary demarcations of events and help to segment the stream of information given instructional videos (van der Meij & van der Meij, 2013).

A recent study by Garret (2020) examined the role of in-built pauses in a software tutorial by comparing a condition with video that included pauses (segmented video) to a condition with video without pauses but comprehensive user control functions to adapt the video pace (unsegmented video). The segmented video stopped automatically at key moments and users were required to manually click to move on to the next video section. In the unsegmented video, user control functions allowed learners to pause, resume and rewind video sections whenever they wanted. Garret's study (2020) showed that contrary to what is advanced by Mayer's segmenting principle (2014), system-based pauses in the video did not reduce learners' cognitive load and also did not improve the learners' performance compared to learners that used unsegmented video and had full control over the video pace. In fact, learners in the latter condition outperformed learners in the segmented video condition. This finding implies that including pauses in software tutorials may not be necessary, as long as users themselves can manage the video pace according to their needs with user control function. All user control functions available to learners in Garrett's study (2020) are by default also available to YouTube users, consequently video designers publishing videos on YouTube do not necessarily need to include pauses in their video design.

YouTube's Chapter Feature

Discussion. A very interesting and important finding with regard to support user control is YouTube's new chapter feature. This feature enables video producers to define different sections in all their videos, irrespectively of when they were first published. When chapters are defined, the video progress bar is divided into different segments with headings. When the users hover over the progress bar, a thumbnail and segment name appear as displayed in Figure 14. At the point at which the videos were sampled as part of the present study, this chapter feature was in a test phase and available only for the YouTube desktop and android version. None of the sampled videos was designed using the chapter feature.

Figure 14

YouTube's Chapter Feature

- *Note*. (source: https://www.youtube.com/watch?v=lTRiuFIWV54).

The chapter feature is interesting for designers and potentially supporting learning for a number of reasons. Visually segmenting the progress bar makes it very easy for users to navigate through the video and consequently supports user control as proposed in guideline three of the 8 G (van der Meij & van der Meij, 2013). In addition, the segmented progress bar signals the video's structure, which is beneficial for learning as outlined in Mayer's signaling principle (2014). Further, when the segments are designed to correspond to a sub-tasks or action steps, they can be seen as checkpoints for the users, which is helpful as they can oversee their own progress in completing the task (Meij & van der Meij, 2013, p.218).

Guideline 4- Preview the task

None of the tutorials in the dataset included a preview. Because the preview was a prerequisite for items 4.1. goal promotion in preview, 4.2. conversational style in preview and 4.3. introduction of new concepts in preview, these items did not apply to any video either.

4.4. Goal illustration

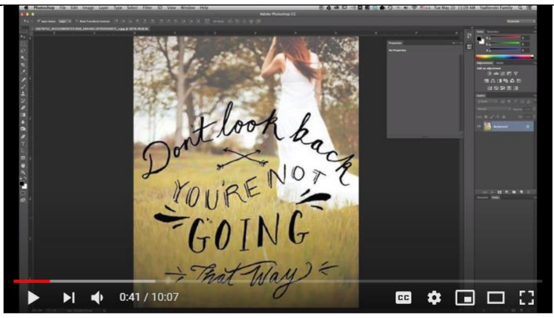
Results. The majority of videos (51,1%) included one type of goal illustration and a further 6,7% of the videos even included two types of goal illustrations. The remaining 42,2% of tutorial did not include any goal illustration. A Kruskal-Wallis test was performed to test if there was a difference in video popularity between tutorials that included one type of goal illustration or several types of goal illustrations. There was no significant difference in video popularity across categories, *H*(2)= 2, 596 p=.273.

Discussion. In total, four different type of goal illustrations were found and will be described below. Most of the tutorials (33%) promoted the goal with an illustration of the design problem, which means that the (often poor) starting-state of a file is explicitly pointed out.

Table ***7*** shows how the design problem was illustrated in an image editing tutorial that had the goal to improve the image quality. The pixelated version of the image was shown while the narrator verbally highlighted several problems arising from this issue. In contrast to the other types of goal illustration, the focus of this type of goal promotion was not on the desired end-result but on the flaws of the file before starting the editing process, which is likely to instigate users to recognise the goal and execute the demonstrated software task.

Table 7-

Illustration of A Design Problem in an Image Editing Tutorial

1. How To Enlarge a Low Resolution Image For Print Using Photoshop Narrator: Hi and welcome to JC Lasky so have you ever had an image that was low resolution it was all pixelated didn't look so good but you wanted to print it out large say an 8 by 10? Well that's the case here with this image. I want to print it out at an 8 by 10 but here even at the size that it's at which is a little over 5 and a half inches by a little less than 7 inches, **it just looks pixelated and it just doesn't look good and it wouldn't reproduce well if we tried to print it as an 8 by 10**. So, a couple steps I would do here to just help make it a cleaner better image. Timing: 0:00-0:37	

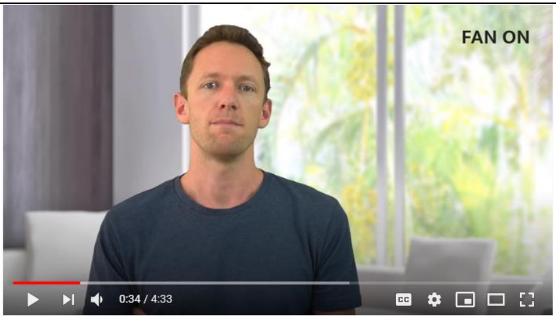
Note. (source: https://Www.Youtube.Com/Watch?V=8u-Lqsvn3fy).

Further, the goal is illustrated by showing before-after displays in 15,6% of the tutorials. In this type of goal illustrations, the instructor presents the start and the goal state simultaneously or subsequently. Before-after displays are expected to be a very enticing and effective type of goal promotion as they combine motivational principles that raise the learners' attention, namely they combine concreteness and the provocation of a menta conflict (Keller, 2010). Table 8 shows how the effect of noise removal in a sound editing tutorial was illustrated by showing a before and after audio-clip. The goal illustration was accompanied by a narrative that mentions commonly occurring problems when recording sound. This narrative clearly promotes the relevance of the task to the users and stresses that no undue effort is necessary to complete this task (Farkas, 1999). To further clarify which audio sample refers to the before and which to the after display, the video producers added on-screen text with state and the name of the noise source.

Table 8

Before-After Display in a Sound-Editing Tutorial

1. How To Remove Background Noise In Videos Narrator: When filming your videos you want to control as many environmental variables as possible background noise can be a huge issue with even relatively quiet noises from devices like air conditioners or fans and louder noises like maybe a computer that's on in the room **they're not only lower the quality of your videos they can also be quite distracting in this video we'll come out easily remove or drastically reduce any constant background noise that might be in your video**. Timing: 0:00-0:28	

2. How To Remove Background Noise In Videos Narrator: to show you what I mean I've got a fan on here in the room with me which isn't something that you'd normally have on while you're filming but it's hot today and this is also a good example for the video so this is the unadjusted audio right now. Timing: 0:29-0:41	
3. How To Remove Background Noise In Videos Narrator: and now future Justin has switched it off in post-production and back on and back off you get the point here's how to do it. Timing: 0:41-0:51	

Note. (source: https://www.youtube.com/watch?v=10FFKl_0GSA).

Third, in 8,9% of the videos the goal was promoted by showing a mock-up illustration of an accomplished task. A mock-up is a simplified version of the desired end-result with exaggerated features, making the effect of the to-be-performed task very clear. Figure 15 shows a mock-up goal illustration for an image editing tutorial. The tutorial goal is to improve the photo resolution, and the animated illustration shows how the photo of an eye changes from a pixelated to a better image quality. This is a mock-up of the actual goal as the tutorial shows how to improve the image resolution on a different, bigger image.

Figure 15-

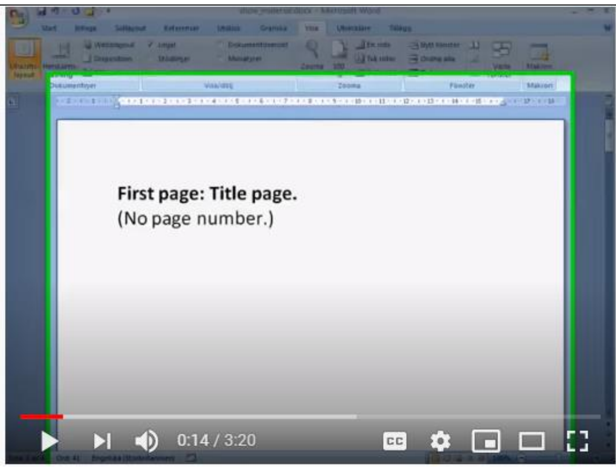
Mock-Up Goal Promotion For An Image Editing Tutorial

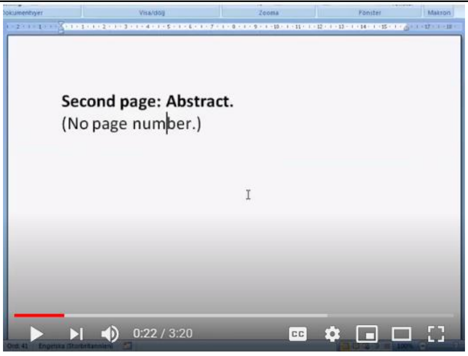
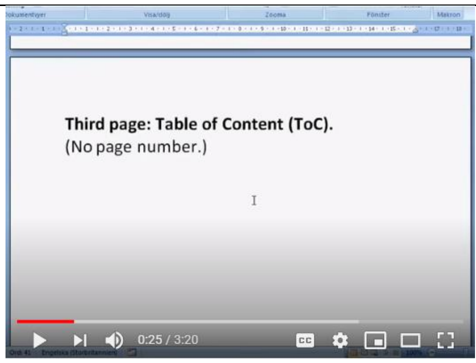
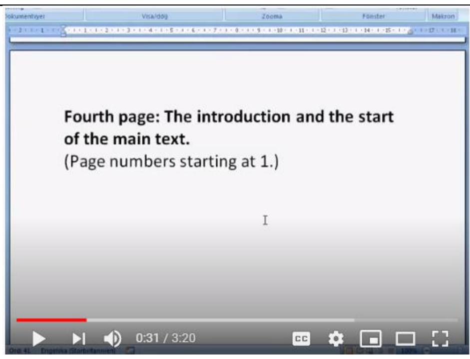
Note. (source:https://www.youtube.com/watch?v=uWAtYPMqPkw).

Table 9 shows the mock-up goal illustration of a text editing tutorial which had the goal to the start page numbering with one, at page number four.The exaggerated feature of this mock-up were the subtitles in parentheses, which highlighted that the first three pages did not contain page numbers. In order to perfectly comply with instructional design principles like Mayer's redundancy principle (2014), the tutorial's goal illustration should not repeat the spoken narration (2005).

Table 9-

Mock-Up Goal Illustration in a Text-Editing Tutorial

1. Word: Page numbers starting from a specific page Narrator: Hello, Word 2007, **page numbers starting from a specific page**. Okay, I am assuming that you would like to create something that looks like this. **You want a first page with no page numbers.** Timing: 0:00-0:18	

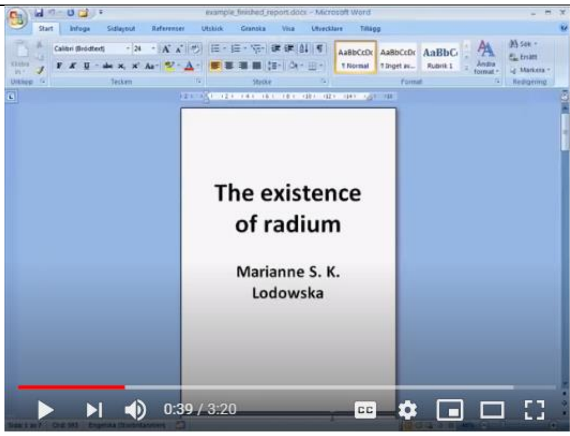
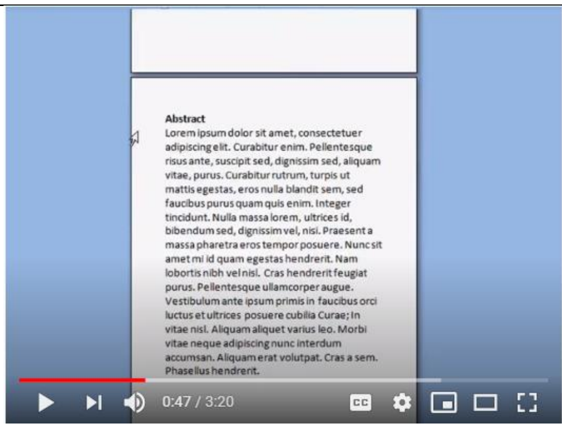
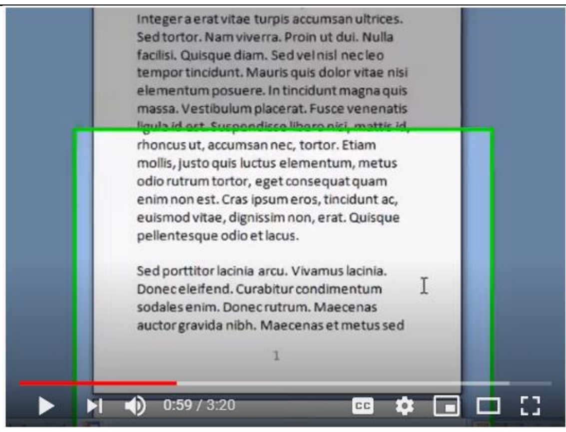
2. Word: Page numbers starting from a specific page Narrator: There are **no page numbers on the second page** Timing: 0:19-0:22	
3. Word: Page numbers starting from a specific page Narrator: or the third page. Timing: 0:23-0:26	
4. Word: Page numbers starting from a specific page Narrator: You want the page number to start here on the fourth page at the beginning of the main text and the introduction. Timing: 0:27-0:34	

Note. (source: https://www.youtube.com/watch?v=NGzz2ZmLrFw).

Fourth, in 6,7% of the tutorials the goal was illustrated by showing the desired end-result. Table 10 displays such a goal illustration in a text-editing tutorial. In this particular tutorial, it is demonstrated how a report was formatted so that page numbers start at page three. In order to promote the tutorial goal to the users, the narrator showed an example report with the desired formatting style at the beginning of the tutorial.

Table 10-

Illustration Of The Desired End-Result In A Text Editing Tutorial

5. Word: Page numbers starting from a specific page Narrator: The finished product should look something like this - a nice title page Timing: 0:35-0:45	
6. Word: Page numbers starting from a specific page Narrator: Then you can toss in an abstract or whatever. Please note that there are no page numbers on these first pages. Timing: 0:46-0:55	
7. Word: Page numbers starting from a specific page Narrator: The page numbers start here on the fourth page and they start with the Arabic number 1. Timing: 00:56- 1:02	

Note. (source: https://www.youtube.com/watch?v=NGzz2ZmLrFw).

Overall, a number of creative and well-designed types of goal promotions were found in the sampled videos and points out potential design issues. This information can hopefully inspire video designers to design a goal promotion that suits their video tutorial.

4.5. Narration style applied in goal illustration.

Results. In 42,2% of tutorials, a goal illustration was accompanied by a conversational narration. Further, 6,7% of the goal illustration were accompanied by a mixed conversational style. The remaining 51,5% of all tutorials did not include any goal illustration, and consequently no verbal goal illustration, at all. A Kruskal-Wallis test revealed that there was no significant difference in video popularity between tutorials that included a goal illustration with a conversational narration and tutorials with another type of narration during the goal illustration, H (2) =2,596 p=.273.

Discussion. Video designers are recommended to accompany the goal illustration with a verbal explanation because it can help to raise the users' perception of task relevance. In addition, as recommended by van der Meij and van der Meij (2013), a goal promotion can be used to stress that the task performance does not require excessive effort, which can motivate users. Also, if the goal promotion uses a conversational style, it is according to Mayer (2014) a social that supports learning from video.

Guideline 5 – Provide procedural rather than conceptual information

Table 11

Descriptive Statistics About Procedural Information Given Dataset

Distribution across the sample		Distribution across tutorial types					
	Total (*N*=36)	Image (N=7)		Sound (N=15)		Text (N=14)	
Mean	0,55	Mean	0,67	Mean	0,41	Mean	0,64
Standard Deviation (SD)	0,27	SD	0,19	SD	0,19	SD	0,32
Range	1,00	Range	0,62	Range	0,70	Range	1,00
Minimum	0,00	Minimum	0,31	Minimum	0,09	Minimum	0,00
Maximum	1,00	Maximum	0,93	Maximum	0,79	Maximum	1,00

Results. In order to check for the extent to which tutorials presented procedural rather than conceptual information, the first 150 words after a potential the goal statement or illustration were reviewed. It was assessed whether information was of procedural, conceptual or mixed (procedural and conceptual) content. Tutorials with a written narration (*N*=7) were deliberately left out of this analysis in order not to skew the analysis because written narrations were found to be extremely short.

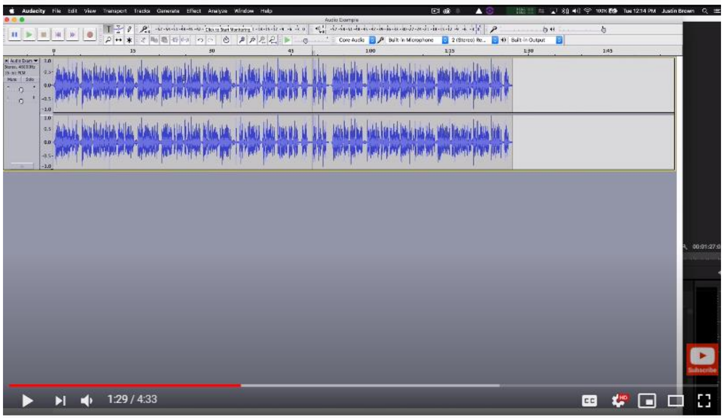
The analysis showed that on average, tutorials with a spoken narration (*N*=36) contained 55% of procedural information and in taken together, 45% conceptual and mixed content. While text and image editing tutorials included about the same percentage of procedural information, 70% and 67% respectively, sound editing tutorials contained the least amount of procedural information of only 41%. Kendall's tau-b correlations were run to determine the relationship between percentage of procedural content included and video popularity. When including all tutorials with a spoken narration (*N*=36), no significant relationship was detected (τb= -.,168, *p*=0,152). However, after excluding the two extreme outliers (*N*=34) a weak, negative correlation between supplying procedural content and video popularity was found (τb= -.,253, *p*=0,036).

Discussion. A potential reason for the high percentage of conceptual and mixed information in sound editing tutorials is that a number of these tutorials included explanation about prerequisites, e.g. which type of software was used and where to download it, as displayed in segment one in Table 12**Fehler! Verweisquelle konnte nicht gefunden werden.**.

Table 12

Transcript of a Sound Editing Tutorial with a High Percentage of Conceptual And Mixed Content

1. How To Remove Background Noise In Videos Narrator: So we're going to use the raw footage off your camera or you can export your video from your video editing software **and the software we're going to use to tidy up the audio is called audacity it's a free program so open that now and I'll put a link in the description below for where you can download it** and will open up your video file audio	

example so this is our audio file **which is actually the video that you can see on the screen behind just make this a bit bigger**. Timing: 0.52-1.28 min	
2. How To Remove Background Noise In Videos Narrator: okay **so what we'll do is we'll just zoom in on our audio timeline here view zoom in view zoom in so what we're looking for is a section like this where we can see that there is some background noise there's no talking the talking's this part here** but when in between when I don't (...) Timing: 1.29-1.52min	

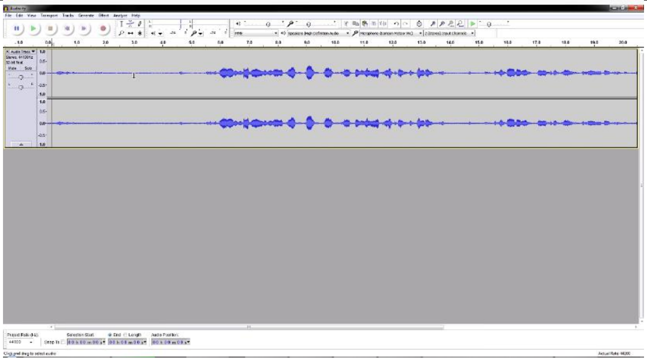
Note. (source: https://www.youtube.com/watch?v=dZ0DWNLrToU).

Another reason for the high percentage of conceptual and mixed content is that the instructors provided background information to make the demonstrated actions and choices comprehensible as shown in segment two of **Fehler! Verweisquelle konnte nicht gefunden werden.**

Furthermore, the particularly high percentage of conceptual and mixed information in the transcript excerpts of sound editing tutorials could stem from the peculiarity of this editing task. The instructors generally show how to manipulate a sound file in order to achieve a certain goal; however, the specific file and problems of the end-user are likely to differ. In order to make the knowledge transferable, the instructors therefore spent more time on explaining how to interpret the spectrograms, how and why settings were chosen and what aspects to bear in mind when editing another file as shown in Table 13.

Table 13

Transcript Of A Sound Editing Tutorial With Explanations Of The Spectogram

1. How to remove background noise with audacity Narrator: Okay let's get started you see here this is a pre-recorded thing I just did it and **if you look at it this is not a flat line this has bumps and curves and thingamajigs and watching knots yeah that's soo man you see those little bumps that is static static static static static static** and that's what you get now to get rid of static in audacity you're going to what I don't wanna do is that either at the beginning or at the end of my videos I will leave about 2 to 3 4 seconds of just white noise not talking not moving not pressing any buttons and I will normally that pick this like this go to effects and go to noise removal and the noise removal dialog box will pop up and says step one and click get noise profile what that does is look at once you select it and whatever you have selected it says this is noise get rid of anything that is like this […] Timing: 0.53-1.58min	

Note. *(source: https://www.youtube.com/watch?v=ZqeG2ZiN_6A).*

The significant negative correlation between supplying procedural content and video popularity indicates that viewers favoured tutorials that supplied more conceptual and mixed content and is surprising as van der Meij and van der Meij (2013) suggest to include only a minimal amount of such information, arguing that viewers' foremost motive for watching software tutorials is their interest in knowing which action steps (i.e. procedural information) to complete to achieve their goal. A possible reason for the negative correlation could be the type of software tasks that the sampled video showed. They could have attracted mainly novice users with limited prior knowledge. This type of learner could appreciate more conceptual information in order to understand the demonstrated actions.

Most importantly, analyses of the transcripts made it abundantly clear that the narrations did not follow the stream-lined step model by Farkas (1999). The streamlined-step model is outlines strategies for the design of effective procedural discourse, that is written and spoken dialogue that supports people in executing tasks. The streamlined step model argues for brevity, simplicity and clarity in procedural discourse (Farkas, 1999): action steps should consist of a brief action statement in the imperative voice, followed by one to two sentences about the system's response to the user action. Optionally, the action step is followed by an example or explanation, always given that it remains brief. Beginning action steps with lengthy descriptive statements violates the streamlined-step model (1999).

These recommendations have frequently been violated in the narrative of the sampled tutorials. As illustrated in

Table 14 and

Table *15*, action sequences were mainly stated using personal pronouns (I, you, we) and using in the gerund or future tense. Compared to using the recommended imperative voice (van der Meij & van der Meij, 2013), this resulted in unnecessarily lengthy phrases and made it difficult to distinguish action steps from the other discourse. This made it difficult to perceive a clear structure in tutorials and made it partly troublesome to follow them.

Table 14

Transcript that violates the streamlined-step model

1. Sony Vegas: Remove Static / Background Noise Tutorial Narrator: So as **you can** slightly here **you can see here like a sort of like** wind and **you know it is** a hissing noise in the background so anyone who was that or any kind of hissing **I'm gonna tell you** how to fix that so what **we need to do** first of all is go on our audio clip **we wanna go** on event FX here make sure **you do not do this on** the video clip and on the audio clip **you want to click that we wanna go to** audio **we're gonna** go on all and express effects audio restoration to look like it and click OK then basically **you want to** set the settings to reduce noise **you want to put** that to about 26 these are just my personal settings which I use 26, 27 kind of thing. […] Timing: 0.51-1.45

Note. (source: https://www.youtube.com/watch?v=_mrsAVSsfG4).

Table 15

Second example of transcripts that violates the streamlined-step model

1. How to Remove Noise From Vocals and Recordings - Edison Method

Narrator:
So **let's go over** to the mixer and take a look at this channel it's an empty Channel and **I'm just gonna** add the Edison to it I've called this D noise **let's select** our waveform and drag and drop it straight into the Edison from here **I'm just going** to enable spectrum and your view and **I'm going to change** the display settings or something that I like the look off now **we're gonna** look for a section that only has the noise it's really important that there's no squeaks or clicks just the noise next we right click on this button to acquire a noise profile and then **we select** the entire region that **we want** the noise to be removed from then **we left click** on the same button again and it will pull up this little clean up tool […].

Timing: 2.05-2.47min

Note. (source: https://www.youtube.com/watch?v=Mf6rjGY8os8).

In sum, the transcript analyses showed that narratives could be significantly reduced in length and improved from an instructional design perspective if a number of principles are applied. By employing the imperative voice for action steps as suggested in Farkas' streamlined step model, reducing verbiage in line with Caroll's minimalism design principles (1990), and removing all unnecessary information from the narrative as proposed in Mayer's coherence principle (2014), the clarity and brevity of the narratives could be substantially improved.

Guideline 6- Make tasks clear and simple

6.1. Description of action sequence

Results. To evaluate if action sequences were described in appropriately, it was checked if three randomly selected action steps with facilitating modifiers per video were stated in their correct order i.e. first stating which interface location to focus on and then the action step itself (Farkas, 1999). Action steps were assessed as incorrectly described in two situations: when the interface location was stated after the action step, e.g. "press print, on the file tab" instead of " on the file tab, press print"; or when imprecise language was used to refer to locations, e.g. "click here" instead of "click on the file tab".

Overall, in 44,4% of tutorials, all three action steps were described in the correct order. One third of action sequences were correctly and precisely described in 11,1% of tutorials. Two thirds of actions sequences were described correctly in 8,9% of tutorials. Further, 2,2% of tutorials included three extremely vaguely described action sequences, which is why they were assessed as incorrectly described. Further, the remaining 33,3% either did not contain a spoken narration or they did not include action sequences with facilitating modifiers, so they could not be analysed with regards to this characteristic. A Mann-Whitney U test revealed that there was no significant difference in video popularity between tutorials that presented all action sequences in their natural sequence and other tutorials, U ($N_{\text{tutorial describing all action sequences correctly}}$=20, $N_{\text{not all action sequenes described correctly}}$ =15)= 87, p=-.572.

Discussion. Overall, the majority of tutorials that included three action sequences with facilitating modifier adhered to the guideline and followed the user's mental plan in describing an action sequence. Table 16 illustrates examples of poorly describes action sequences: the narrator's verbal descriptions are imprecise because steps are left out and exact names of menu options selected are not stated. In addition, the action sequences were executed at a quick pace and the user interface was not always easily legible. Consequently, it was very burdensome to follow the video, and the video can cause type I essential overload. Type I essential overload describes a situation in which both the learner's visual-pictorial and auditory-verbal channels are overloaded. As a result, the learner may be unable to perform the cognitive processing necessary for comprehending the presented information, which hampers learning (Mayer, 2014).

The description of action sequences is ideally short and simple (Farkas, 1999). However, the analyses showed that the descriptions of action sequences were often lengthy because the imperative voice was not employed and because action sequences included superfluous words (Table 17). This finding again stresses the need to painstakingly design the narrative.

Table 16

Vaguely Described Action Sequences in A Sound Editing Tutorial

1. How to Remove Noise From Vocals and Recordings - Edison Method Narrator: Let's go over to the mixer and take a look at this channel. It is an empty channel and I'm just going to add the Edison to it. **[actually performed steps: selects slot 1, then under select, goes to Misc, and finally selects Edison]** Timing: 2.04-2.09	
2. How to Remove Noise From Vocals and Recordings - Edison Method Narrator: From here I'm just going to enable spectrum, and dual view, and I'm gonna change the display settings to something that I like the look of. **[actually performed steps: in Edison menu, clicks on eye icon, selects spectrum, dual view, and then in display settings under spectrum gradient, selects Glacier.]** Timing: 2.18-2.23	
3. How to Remove Noise From Vocals and Recordings - Edison Method Narrator: Now simply select and drop and drag this cleaned up audio onto the playlist. **[actually performed step: selects relevant section from spectrogram, then in Edison menu selects**	

mouse cursor icon, drags selection to playlist, then closes file menu.] Timing: 3.35-3.42	
4. How to Remove Noise From Vocals and Recordings - Edison Method Narrator: Rename It something useful, give it a nice colour and send it somewhere on the mixer. **[actually performed step: somehow opens up new menu and file name, renames it, clicks on triangle icon next to name to change colour, closes menu and somehow sends it onto the mixer.]** Timing: 3.43-3.47	

Note. (source: https://www.youtube.com/watch?v=Mf6rjGY8os8).

Table 17

Examples of Poorly Worded Action Sequences

▪ So, what **we'll** do is **we'll** just zoom in on our audio timeline here view, zoom in. ▪ **Now we come up** here to effect, and choose noise reduction. ▪ **We go back** up to effect, and **we choose** noise reduction. (source: https://www.youtube.com/watch?v=dZ0DWNLrToU)
▪ Then **we'll** come up to the top to file, new from clipboard. ▪ […] to do this **we'll come** up to the top to document, resize document in this dialog box. ▪ Then **we'll** going to the live filter icon, and apply a high pass filter. (source: https://www.youtube.com/watch?v=K1DoEPYosJ0)
▪ Go effect and **I'm just going to** type in adapt. There it is adaptive noise reduction, **I can drag and drop it on this audio clip**. ▪ **We're gonna** go window, audio track mixer.

- Hit the drop down and **I can choose** noise reduction/restoration.

(source: https://www.youtube.com/watch?v=ku6ZZ8yhuls)

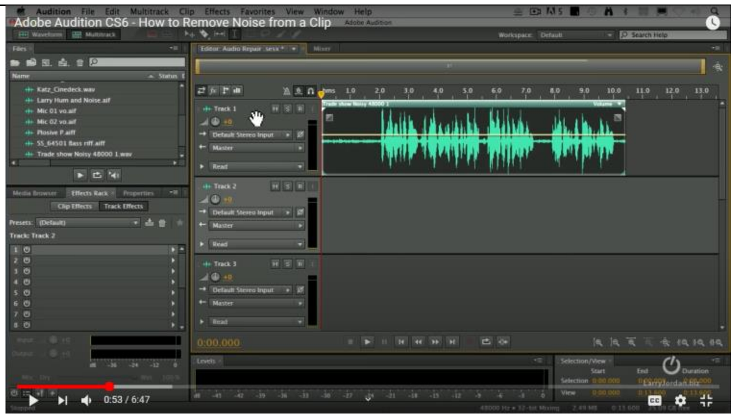
6.2. Highlighting the interconnection of the user action and system reaction.

Results. The interconnection between the user's actions and the system's response was assessed highlighted when an explanatory narrative of what task can be done was coupled with the visual demonstration of how this was achieved (van der Meij & van der Meij, 2013). This was measured with a 7-point Likert scale ranging from 'never' to 'very frequently'. A spoken narration was required for assessment of this design feature because there is a high chance that users may not be able to spot the system reaction while reading on-screen text. In 20% of tutorials, there was no spoken narration, therefore they were excluded from this analysis. The majority of tutorials they highlighted the interconnections frequently (15,6%) or very frequently (62,2 %). Only 2,2% of videos only occasionally highlighted the interconnections. A Kruskal-Wallis test indicated that there is no significant relationship between following the user's mental plan in describing action sequences and video popularity, H (2)=3,434 p=.180.

Discussion. Generally, the software reacts to the user's actions, and this reaction is a form of immediate feedback (van der Meij & van der Meij, 2013). If the action was executed correctly, the system will react correctly. If users have been introduced to the desired software reaction and understand them, they can assess whether they have executed action(s) correctly or not. Therefore, the software demonstration should stress both the user's action and highlight the system's reaction. An example of very well described interconnection is displayed in Table 18. In this example, the narrator precisely explains the visual changes in the user-interface that are caused by the user selecting a menu option.

Table 18

Well Designed Verbal Highlighting Of The User Action And The System Reaction (

1. Adobe Audition CS6 - How to Remove Noise from a Clip Narrator: We can get rid of this one is in real time in the timeline and the other is permanently in the waveform editor listen to this effect here's how it works you select the track **notice that as I click on the track light grey indicates the track is selected dark grey says it is not**. Timing: 0.39-0.55	

Note. (*source: https://www.youtube.com/watch?v=Y-_JGy6fWeY).*

In one tutorial (2,2%) the interconnections were highlighted only occasionally and thus insufficiently. The narrator introduced a number of actions that were actually not demonstrated. Also, a series of still screenshots were used in the beginning of the tutorial, which inhibited the users see the sequence and relation of user actions' and the system's response.

6.3. Highlighting

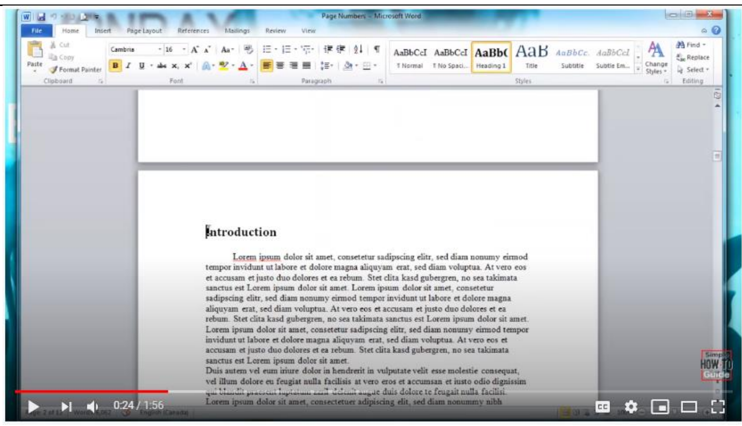
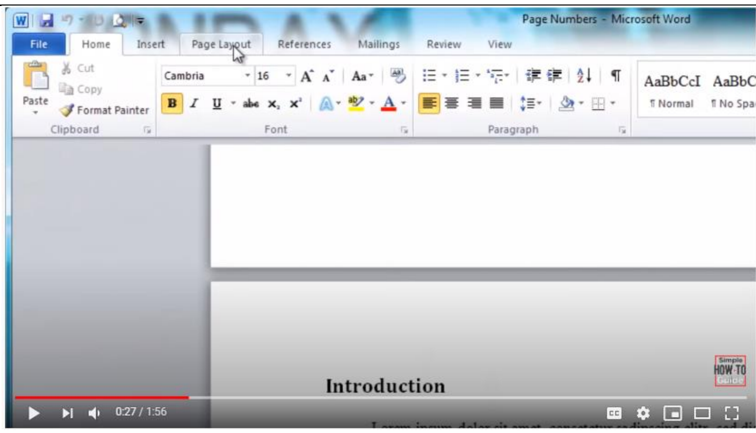
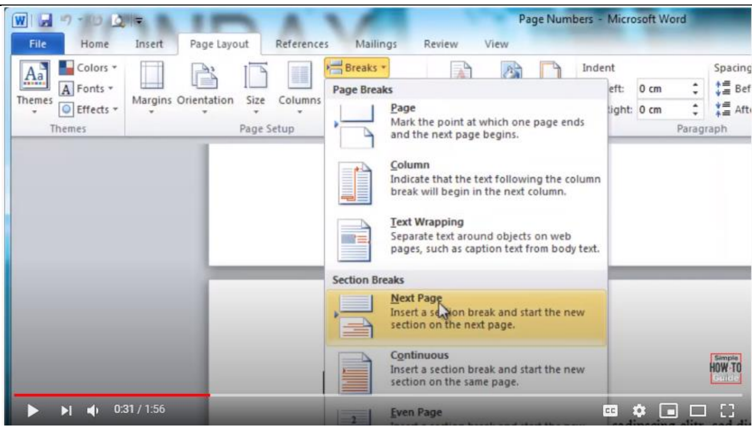
Results. In the majority of tutorials, at least one type of highlighting was applied (42,2%) and in a considerable number of tutorials, two to three types of highlighting were applied (22,2%). In contrast, in 33,3% of the tutorials, no highlighting was applied at all. A Kruskal-Wallis test revealed that there was no significant difference in video popularity between tutorials that apply any one type of signaling, tutorials that apply several types of signaling, and tutorials that do not apply signaling, $H(2)=2{,}596$ $p=.273$.

Discussion. Signaling is very similar to what De Koning, Tabbers, Rikers and Paas (2009) described as cueing, which can be classified into serving three functions: selection, these cues that guide the viewers' attention to particular locations; organisation, these are cues that indicate structure; and integration, these are cues that illustrate the relation within and between elements. When used, highlighting was predominantly used to guide the users' attention to specific screen locations and in line with van der Meij & van der Meij's recommendation, the cues were always clearly imposed and not to be confused with elements of the user interface (2013). In total, five diverse types of signaling were found in the sample and will be illustrated in the following.

To start with, the most commonly used highlighting technique were zooms (40%). As exemplified in Table 19, zooms were applied to guide the users' attention to relevant screen elements such as specific menu tabs and to support the legibility of the user interface.

Table 19

Highlighting Technique Zooms

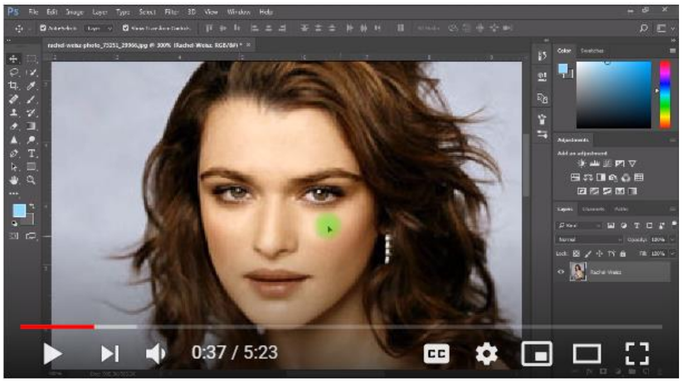
1. Page Numbers Starting at a Specific Page in Word 2010 Narrator: All you need to do for this is scroll down to that page that you want the numbering to start on. Timing: 0.15-0.23	
2. Page Numbers Starting at a Specific Page in Word 2010 Narrator: And you're just going to click rate at the very beginning of the page. Timing: 0.26-0.28	
3. Page Numbers Starting at a Specific Page in Word 2010 Narrator: Then go to the page **Layout tab and then over to breaks and then down to the next page**. Timing: 0.29-0.31	

Note. (source: https://www.youtube.com/watch?v=PHBQIwaC9ts).

The second most commonly used highlighting type was signaling of the mouse cursor (31,1%). The mouse cursor was highlighted with shades in different colours (compare Figure 16, Figure 17, Figure 18), by enlarging it (Figure 17), or by adding animated circles around it whenever a mouse click was performed (Figure 19). While adding animated circles around the cursor was perceived as rather distracting and less helpful in guiding attention, shading the mouse cursor improved its visibility especially in tutorials recorded in full screen mode or tutorials that showed dark and complex user interfaces such as the ones displayed in Figure 16. The shading helped the viewers to locate and follow the cursors more easily and to better see which specific menus and options were selected.

Figure 16

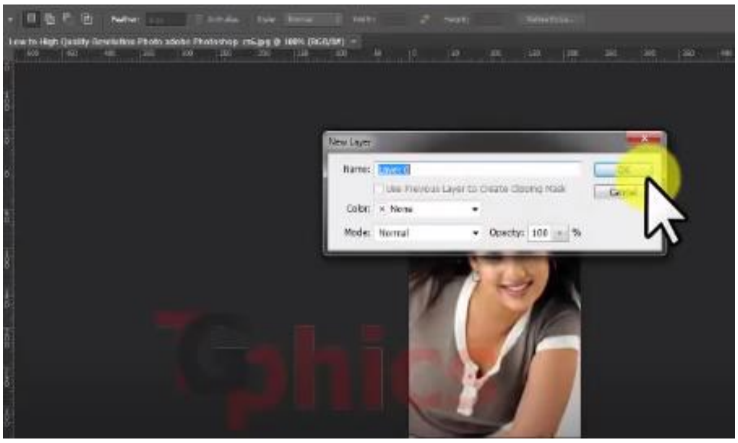
Shading of The Mouse Cursor in A Dark, Detailed User-Interface



Note. (source: https://www.youtube.com/watch?v=uWAtYPMqPkw).

Figure 17

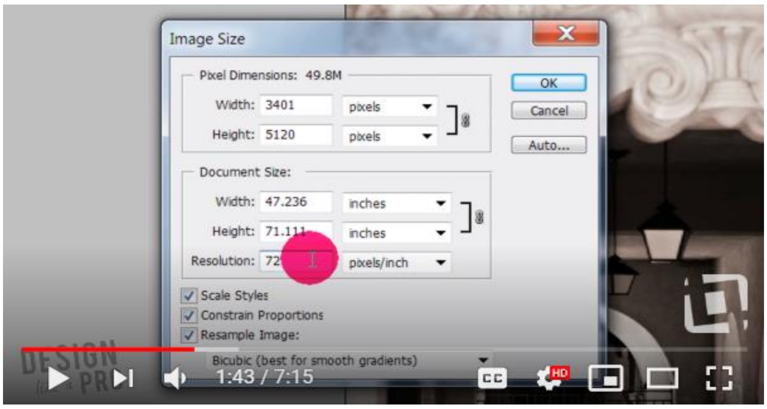
Enlarged, Shaded Mouse Cursor



Note. (source: https://www.youtube.com/watch?v=95QkksEQiME).

Figure 18

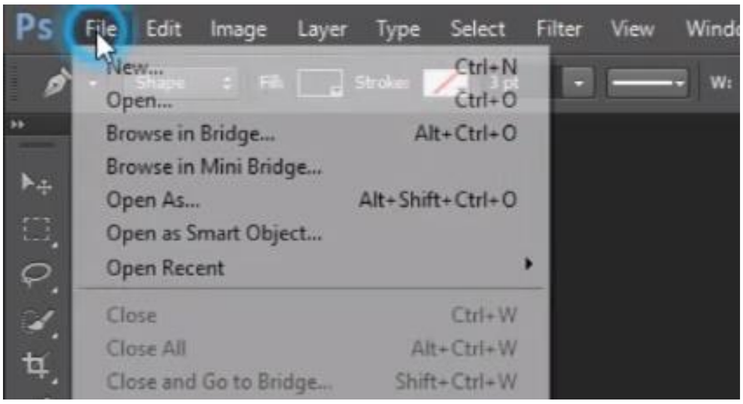
Shaded Mouse Cursor



Note. (source: https://www.youtube.com/watch?v=K4LRaFxZZvI).

Figure 19-

Animated Mouse Cursor

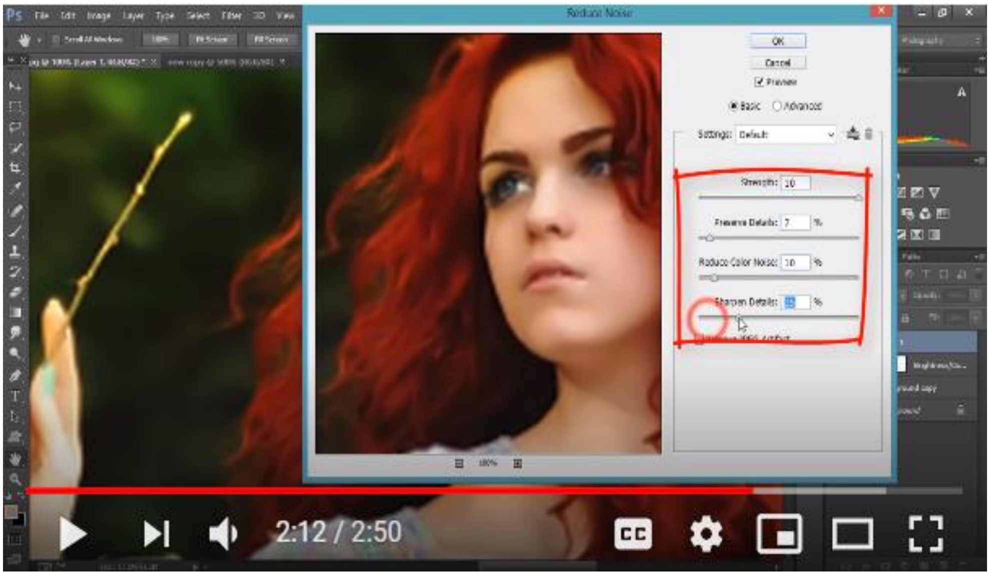


Note. (source: https://www.youtube.com/watch?v=UcElYhmeutw).

In 11,1% of tutorials, frames were used, predominantly to emphasize which menu items to select or which settings to adapt as displayed in Figure 20. It is noteworthy that in four out of the five videos that employed frames, a spoken narration was lacking. In these cases, frames were the non-verbal method to transmit procedural information.

Figure 20-

Frames Used To Highlight Menu Option

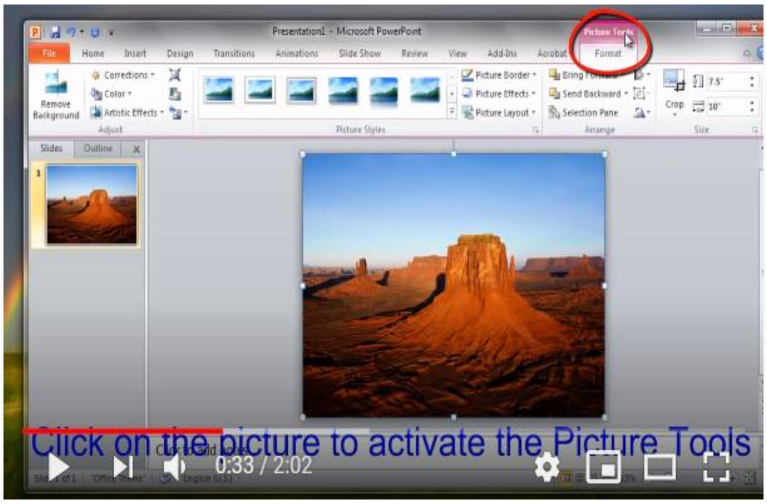


Note. (source: https://www.youtube.com/watch?v=K4LRaFxZZvI).

Figure 21 illustrates the central issue of tutorials that use on-screen text: users face the problem of having to divide their visual attention between the written text and the actually relevant elements of the screen. Also, given the transient nature of video, users might not be quick enough to read the text and spot these elements at the same time, which can hinder their learning.

Figure 21

Frames Used To Signal A Menu Tab In A Tutorial With On-Screen Text

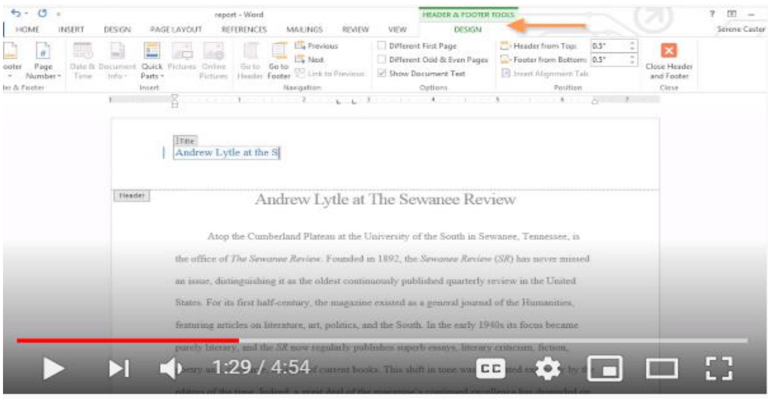


Note. (source: https://www.youtube.com/watch?v=NFGx78yFvy8)

In 8,9% of the tutorials, arrows were employed to highlight menu tabs (Figure 22), icons (Figure 23), and relevant details in the task process (Figure 24).

Figure 22

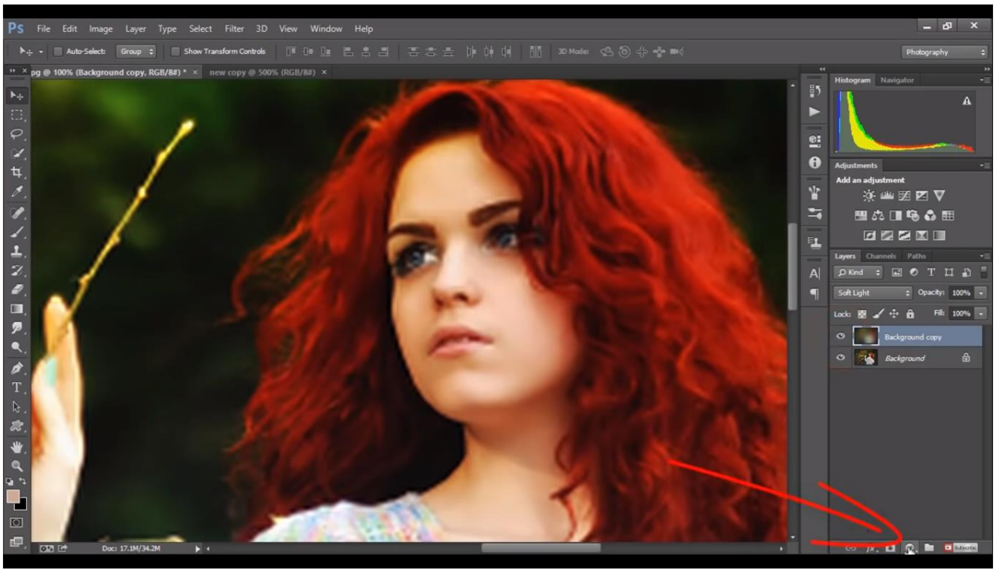
Arrow Highlighting A Menu Tab



Note. (source: https://www.youtube.com/watch?v=BU6tOwzWtY8).

Figure 23

Arrow Indicating Small Menu Icon



Note. (source:source: https://www.youtube.com/watch?v=K4LRaFxZZvI).

Figure 24-

Arrow Signaling Important Details In Editing Process

Note. (source: https://www.youtube.com/watch?v=Z1CyFNoAWZc).

Finally, there was one tutorial (2,2%) in which mixed type of signaling was used. As depicted in Figure 25, this type of signaling resembled a magnifying glass and was used zoom into a screen element. Unfortunately, the instructor moved this magnifying glass rather quickly and it overshadowed other elements of the screen, thus it was not perceived as an ideal application of signaling.

Figure 25

Mixed Signal Type

Note. (source:https://www.youtube.com/watch?v=Y-_JGy6fWeY).

Overall, occasional zooms into the user interface were perceived as the most suitable method to increase legibility of the screen elements. In addition, shading the mouse cursor with bright colours was another signaling method that served its goal well to guide the viewers' attention. Contrary, using arrows or frames to highlight screen elements might have led the users' attention well, but did not improve legibility of the signaled elements and was therefore found to be less optimal. Future research could investigate which type of the found signaling cues is most effective in guiding the viewers attention, so that video designers can make even better-informed choices when selecting cues.

Guideline 7- Keep videos short

7.1. Video length

Results. Overall, the average video length was 5:43 minutes (range 1:50 – 14:01 min, *SD* 2.76). A Kendall's tau-b correlation was run to determine the relationship between video length and video popularity amongst the 45 sampled videos. No significant relationship between video length and video popularity was detected (τ_b=0,08, *p*=0,44).

Discussion. With almost 6 min, the sampled videos are quite long bearing in mind to keep the videos as short as possible as recommended by van der Meij & van der Meij (2013). Also, the videos of this study are considerably longer than the videos for declarative knowledge development found by Ten Hove and van der Meij (2015), which were on average 3:35 min. In a study investigating the user behaviour during 6.9 million MOOC video watching sessions, Guo et al. (2014) found that users watch most of shorter videos, i.e. under 3min, and at maximum engage and watch video for 6min in one sessin. Therefore, Guo et al.(1014) recommended 6 minutes as the maximum video length for videos and suggest chunking longer videos into 6 min segments, if needed. This stresses the importance of keeping videos short once more as videos can only support learning if they are watched- so video designers are adviced to bear a maximum minute of 6 min, and an ideal length of under 3 min in mind to increase watch time. Equally, a shorter video length or the split into several videos also contributes to Mayer's segmenting principle effect, which holds that learning is enhanced when learners are presented with manageable amount of information (2014).

Table 20

Video Length In The Dataset In Minutes (Min)

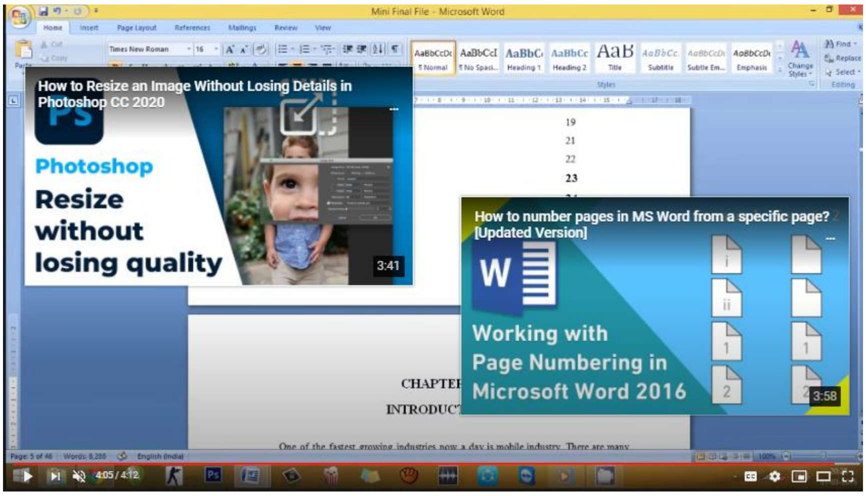
Average video length in %		Average video length per tutorial type					
	Total (*N*=45)	Image (*N*=15)		Sound (*N*=15)		Text (*N*=15)	
Mean	5:43	Mean	4:77	Mean	5:33	Mean	4:16
Standard Deviation (SD)	2:76	SD	3:42	SD	2:38	SD	1:64
Range	12:51	Range	12:51	Range	9:59	Range	4:73
Minimum	1:56	Minimum	1:56	Minimum	2:57	Minimum	1:56
Maximum	14:01	Maximum	14:01	Maximum	12:16	Maximum	6:29

7.2. Show one rounded off task

Results. The majority of tutorials showed only one rounded-off task (62,2%), however, 35,6% of tutorials showed how to execute one editing task on several examples (e.g. multiple images or sound files) or showed several tasks (e.g. in an image editing tutorial, showing how to improve the image resolution and how to sharpen it). One tutorial (2,2%) did not show a clearly rounded off task because the producer placed screenshots with hyperlinks so prominently on the screen that the task demonstration was not visible anymore (Note.). No significant relationship between video popularity and showing one rounded off task was found, (U=198, p=0,349; $N_{\textit{one rounded off task}}$ =28, mean rank=21,57; $N_{\textit{task not rounded off or more than one task shown}}$= 17, mean rank= 25,35).

Figure 26

Poorly Placed Hyperlinks



Note. (source: https://www.youtube.com/watch?v=-px-u-8aCGI).

Disucssion. In line with Mayer's segmenting principle (2014), van der Meij and van der Meij (2013) recommend video designers to presenting the users with "meaningful and manageable units of information" (p. 222). The authors suggest that video designers may have to split videos into several tutorials by creating useful sub-tasks, that its task that make sense to be fulfilled on their own irrespective of the other videos shown. This is recommendation seems useful with regards to the sampled tutorials that showed how to complete several tasks or examples.

Van der Meij and van der Meij (2013) also suggest splitting the videos into smaller sub-tasks and to display them with a table of contents with discernible segments based on the notion that the tutorials would be published on an independent website that allows for such a design. This is not possible on YouTube. Given the new chapter feature however, video designers can give videos a meaningful structure by splitting the video into segments based on sub-tasks and consequently, they may not have to split them into seperate tutorials.

Mayer's segmenting principle (2014) is operationalised as video that automatically stops at key points and that includes a buttons which has to be manually pressed to continue video play. This type of segmenting has not been found in any of the sampled videos. This raises the questions on how realistic and useful this operationalisation and the principle is in the context instructional videos that are used by actual learners, and not only in experiments (Garrett, 2019). This calls for a new operationalisation of the segmenting principle. Also, research into the optimal overall video length and the effectiveness of segmenting video with the new video chaper feature could be helpful to support video designers in the construction of effective software tutorials.

Guideline 8- Background music

Results. The majority of software tutorials did not include music (68,9%) but in contrast, in 31,1% of tutorials music was audible during the demonstration of tasks. A Mann-Whitney U test indicated that there was no difference in video popularity across videos that included or did not include background music during the task demonstration, U ($N_{no\ \text{background music}}$ =31, $N_{background\ music\ included}$ =14) = 154,00, p=.122).

Discussion. Mostly, the background music clearly played a minimal role. In one text editing tutorial however, the background music was very loud and disturbing, which made it difficult to clearly understand and follow the spoken narration. In a sound editing tutorial, a very soft background music was added to the task demonstration. While the music did not compete with the narration in this particular case, the background music ran counter to the point of the video which was to demonstrate how well a software feature removed noise from a sound file. In the other tutorials, the music was not perceived as disturbing

So far, research on background music in multimedia material is inconclusive. A recent literature review on the effects of background music on learning by de la Mora Velasco and Hirumi (2020) showed that between 2008 and 2018, only three studies addressed the topic of background music in multimedia learning modalities. The studies were conducted the effect of background music in digital video games (Linek et al., 2011), immersive virtual worlds (Richards et. al, 2008) and interactive CD-based language learning. Contrary to the cognitive load theory and the multimedia design theory, which discourage the use of background music (Mayer, 2014), these studies found that background music had a positive impact on learning: they reported an increase in student motivation (Linek et. al, 2011), improved recall of facts (Richards et al., 2008) and enhanced foreign language learning (Kang and Williamson, 2014). While these findings are promising, they do not focus on video software tutorials, so findings may not be translatable. Further research is needed to show whether background music can have positive impact on learning when applied to video software tutorials. As pointed out by de la Mora Velasco and Hirumi (2020) further systematic research is necessary and should consider factors such as the characteristics of the learners and the specific music type to paint a clearer picture on the issue. Overall, because the background music in software tutorials does not constitute information that is needed to achieve the instructional objective, music is assessed as extraneous load which should generally be avoided (Sweller, Ayres, & Kalyuga, 2011).

Video designers are therefore advised to not employ background music during the task demonstration.

Conclusion

In the following section, the main conclusions of this study are presented concerning the main research questions.

Characteristics of YouTube's Instructional Videos for Software Training

The first goal of this study was to identify instructional design characteristics of popular software tutorials on YouTube by comparing them against proven guidelines for the design of effective instructional videos. To do so, an analysis framework was developed, which is based on van der Meij and van der Meij Eight Guidelines for the Design of Instructional Videos for Software Training (2013) and further relevant instructional design characteristics known from literature. The analysis showed that the majority of tutorials adhered to most of the eight guidelines as highlighted with ticks in Figure 21.

Figure 21

Framework for Analysis of Instructional Videos, based on van der Meij & van der Meij (2013).

Guideline 1: Provide easy access

1.1: Title structure

1.2: Tutorial goal(s) and title match ✓

1.3: Title specifies context

1.4: Title specifies condition ✓

Guideline 2: Use animation with narration

2.1: Actual interface shown ✓

2.2: Narration style

2.3: Narrator visibility

Guideline 3 Enable functional interactivity

3.1: Video pace ✓

3.2: Pauses in narration

3.3: Enable user control ✓

Guideline 4: Preview the task

4.0: Preview

4.1: Preview includes goal promotion

4.2: Conversational style in preview

4.3. Introduce new concepts ✓

4.4: Goal illustration

4.5: Conversational style in goal illustration ✓

Guideline 5: Provide procedural rather than conceptual information

5.0: Percentage of procedural content ✓

Guideline 6: Make tasks clear and simple

6.1: Description of action sequences ✓

6.2: Interconnections of user's action & system's reaction highlighted ✓

6.3: Highlighting techniques ✓

Guideline 7: Keep videos short

7.1: Video length

7.2: One rounded off task shown ✓

Guideline 8: Music

8.0: Absence of background music ✓

Tutorials title and goals matched; the actual user interface was shown and a human voice for a spoken narration was used; tutorials had an adequate video pace based on the use of a conversational narration speed; further, despite none of the tutorials including a preview, the majority promoted the tutorial goal with an illustrative example and used a conversational narration style to enhance perceptions of task relevance in their goal promotion; in addition, an analysis of the transcript excerpts indicated that, overall, the tutorials provided more procedural than conceptual information. Also, action sequences with a facilitating modifier were mainly described in line with the user's mental plan. Moreover, the interconnection of the user's action and the software's response were highlighted and the majority of tutorials applied signaling techniques to direct the users' attention to relevant screen elements. Finally, YouTube's standard video player provided extensive user control functions by default and thus supported viewers in becoming active learners and adapted the video playback to their processing capacities.

In contrast, there was a number of principles that the majority of tutorials generally did not comply with. To start with, tutorials did not employ the recommended title phrasing advice and instead mainly had a how-to title structure. Additionally, none of the tutorials included a task preview or introduced relevant new concepts in the preview. Moreover, tutorials were quite lengthy overall. In sum, the in-depth video analyses showed which instructional design characteristics were used more frequently and those which were used less in popular software tutorials. Furthermore, this paper provides useful positive and negative design examples of the guidelines and provides explanations as to why poorly applied design characteristics can be detrimental for learning.

The Relation of Video Characteristics to Video Popularity

The second of research question was how the video characteristics that were included in the framework affected video popularity. It was found that only two video characteristics significantly impacted video popularity: showing the narrator's face positively influenced video popularity, while including more procedural knowledge negativeley influenced video popularity. This study thus made has an important first step into understanding which instructional design characteristics of video software tutorials influence video popularity. The results stress importance to continue the search for characteristics that do influence video popularity in YouTube's popular instructional videos for software training.

Video Characteristics To Include To Support Video Popularity

The statistical analyses showed that showing the narrator either permanently as a picture-in-picture montage or as full-screen in task-demonstration-free scenes positively related to video popularity. There was no significant difference in video popularity between the two alternatives. Obtaining an increased video popularity when showing the narrator is in line with earlier findings by Guo et. al (2014) and could stem from an increased social connection between the instructor and the viewer (Fiorella & Mayer, 2018). With regard to the effect of a visible narrator on learning with instructional videos, research has been limited and inconclusive so far. While Guo et. al (2014) found an increased video engagement and attempts at problem solving in MOOCS videos showing the instructor at adequate moments, Wemeskerken, Ravensbergen and van Gog (2018) did not find any difference in learning outcomes by comparing student groups that were shown a video that constantly showed the narrator's face, compared to a control group that was shown a video without a visible narrator. The authors point out that a human face is a strong visual cue that caught the attention of learners (Wemeskerken, Ravensbergen & van Gog, 2018). Constantly displaying the instructor thus bears the risk of being too distracting for learners, meaning that risks and benefits of showing the instructor's face need to be carefully considered. In sum, the recommendation for video designers is to include the narrator's face only in scenes in which the viewers do not have to visually focus on any other aspects of the video; opportune moments are thus the beginning or ending scenes of tutorials.

The second significant relationship of a video characteristic to video popularity was surprising: a weak negative correlation between video popularity and the percentage of procedural information provided in the narration was found. This suggests that viewers preferred videos that contained more conceptual information, which is at variance with recommendations by van der Meij and van der Meij (2013) to keep this type of information to a minimum - the authors expect viewers of software tutorials to be mainly interested in the procedural information. Apparently, this was not the case and needs to be further explored to be fully understood. Interestingly Chong (2018) found in her analysis of popular makeup and hair tutorial that procedural discourse was concerned a number of other topics than demonstration, doing and explaining. Chong suggests to consider including humour and personal narratives, as these have been covered in the popular videos she analysed (2018). Further research is needed to shed light on which content in narratives influences video popularity and why. Based on the findings of this study, video designers could include a significant amount of conceptual and other information as users apparently dislike tutorials that mainly include procedural information.

Proposal of Video Characteristics That Support Theoretical Effectivness of Instructional Videos For Software Training

The literature review and video analyses also yielded a number of l findings that are highly relevant for instructional video designers. In line with research and literature, a number of guidelines that can theoretically support the effectiveness of instuctional videos for software training are proposed.

To begin with, the video analyses revealed how intertwined the guidelines suggested by van der Meij and van der Meij (2013) are with a spoken narration, while it is never explicitly stated as such. The narration relates to the following aspects: narration pace, goal promotion, the provision of procedural rather than conceptual information with the narration, the description of action sequences and the highlighting of the interconnections of the user's actions with the system's reaction. Tutorials that included no narration or only a written one were not able to comply with said guidelines and even the tutorials that included a spoken narration showed many flaws and a considerable potential for improvement.

The review of the transcripts and the analyses of the video characteristics demonstrated issues in the design of many narratives. Action steps were frequently stated using the personal pronouns “I”, “you”, “we” and the gerund or future tense was used, causing the narrations to be overly wordy and complex. This made it difficult to clearly discern action steps from the discourse and contributed to a perceived lack of structure in the tutorials. Consequently, narration did not adhere to the recommendations to cut out all superfluous words and to use the imperative voice for describing actions steps (Farkas, 1999; Mayer, 2014; van der Meij & van der Meij, 2013). This meant that they also did not comply with the streamlined-step model proposed by Farkas (1999). Improving the narrations in line with this model would overall make tasks more clear and easy (guideline 6), which means that the instructional material would induces less extraneous load for learners, which is beneficial for learning.

Moreover, the literature review showed that more concrete design advice for how to pace the narrative is now available. Research by Miao et. al (2018) indicates that the learners’ age and academic background are interacting with each other and have an important impact on the most appropriate speaking rate for instructional videos (guideline 3). More specifically, they found that older learners prefer slower narratives and that instructional videos for professional learners (i.e. learners with extensive prior knowledge about the topic) from an ample age group are best designed with a higher rate of speaking of 210-240 wpm. In contrast, instructional videos directed at amateur learners (i.e. novices with no prior topic knowledge) are effective when a wider speaking rate range from 179-240 wpm is applied. Instructional video designers are thus advised to carefully consider their target groups academic background and age when pacing the tutorial.

Further, more information on the relevance of segmenting is available based on Garrett’s study on Mayer’s renown segmenting principles (2020). Mayer’s principle advocates for providing the user with meaningful and manageable amount of information (2014). So far, segmenting has been operationalised as video that automatically stops at key moments and requires users to press play in order to proceed (Mayer, 2005b; de Koning et. al, 2018). When comparing such a segmented software tutorial to an unsegmented tutorial which offered user control functions to pause and rewind, Garret found that segmentation did not reduce cognitive load and also did not improve transfer (2020). Learners were able to manage their own cognitive load by pausing and rewinding the video (Garret, 2020). For video designers, this suggests that videos do not need to include system-based pauses when users have said control functions. This finding also calls for a re-assessment and new operationalisation of Mayer’s segmenting principle that does not depend on the users to act on an automatically stopped video - a design characteristic that appears artificial and unusual outside of educational research.

In addition, Garret’s findings (2020) connect very well with the chapter feature of the standard YouTube player that was introduced in April 2020. Instead of having to rely on system-based pauses for segmenting a video, YouTube channel owners can now define separate segments in the video, which results in the video progress bar being split into visible segments. To fully exploit the new function in line with effective instructional design characteristics, producers should define video segments based on the individual action steps or subtasks that need to be completed to achieve the tutorial goal. Additionally, these segments can then be titled, e.g. introduction, step 1, step 2 etc. If producers do so, bearing in mind to keep each video segment meaningful and manageable in terms of presented information, this aligns with guideline 7 by van der Meij and van der Meij (2013) and supports Mayer's definition- but not the prevalent operationalisation- of the segmenting principle (2005b). Further, segmenting the video progress bar facilitates the user's navigation through the video and thus its usability (guideline 3, van der Meij & van der Meij, 2013). Additionally, it supports Mayer's signaling principle as it highlights the video structure visually.

A summary of the presented recommendations for video designers is presented in Figure 27. These guidelines are based on the 8G by van der Meij and van der Meij (2013) and have been adapted and extended with novel items that are marked in green.

Figure 28

Updated Guidelines For The Design Of Video Software Tutorials based on Van Der Meij & Van Der Meij (2013)

Guideline 1: Provide easy access	**Guideline 4: Preview the task**
1.1: Craft the title carefully	4.1: Promote the goal
1.2: Specify the software name & edition in the title	4.2: Use a conversational style to enhance perceptions of task relevance
Guideline 2: Use animation with narration	**Guideline 5: Provide Procedural rather than conceptual information**
2.1: Be faithful to the actual interface in the animation	
2.2: Use a spoken human voice for the narration	**Guideline 6: Make tasks clear and simple**
2.3: Carefully design the narration in line with the streamlined-step model by Farkas (1999)	6.1: Follow the user's mental plan in describing an action sequence
2.3: Show the narrator's face in opportune moments	6.2: Draw attention to the interconnection of user actions and systems reactions
	6.3: Use highlighting to guide attention
Guideline 3 Enable functional interactivity	**Guideline 7: Keep videos short**
3.1: Pace the narration according to the learners' age and academic background	7.1: Demonstrate how the tutorial goal is achieved based on one clear exemplary task
3.2: Enable user control	
3.3: Define video segments, divide the video progress bar accordingly and label the segments segments	

Limitations of the Present Study

This study has limitations that need to be acknowledged. First of all, it is important to point out that the used sample size of 45 videos does not allow for robust significance testing. Further, the relative proportions in this study should only be seen as an indication of the uncovered patterns. This study analysed the applied guidelines in a fine-grained manner and not all discovered characteristics were present in sufficient numbers to allow statistical testing; this refers also to the to the category 'narrator visibility in scenes without task demonstration' as this characteristic was found only four times.

Second, the focus of this study was on popular tutorials. Video provides us with information about what viewers appreciate. However, what learners appreciate is not always best for their learning. Salomon (1984) found that learners put less mental effort into learning with a medium which they report to like, and finally learned less with the preferred medium. Salomon (1984) hypothesised this was caused by learners' belief that learning with a medium they like means that learning with this medium is easier and less effort needs to be made to learn.If viewers of instructional YouTube videos have similar views and also make less effort in learning, this may be detrimental for their learning. learnig t Video characteristics that users favour may therefore not necessarily have a positive impact on their learning and each video characteristics should be rigourously tested for its influence on learning.

Third, while the study applied a formula for video popularity assessment that is based on an investigation into viewer rating behaviour (Shoufan, 2019), more research is needed to understand the viewers' motives for not leaving any appraisal in the form of likes or dislikes. Overall, only a small number of viewers (less than 0,01%) left ratings in comparison to the total number of views. As of now it is unclear if not leaving any appraisal is a sign of appreciation or that of dissatisfaction with the video. In addition, it is not entirely clear how YouTube counts the number of views, if for instance watching the video for only a few seconds is enough for YouTube to count as one full view. Therefore, the number of views may have been overestimated in this study. More research and new insights on these issues and may lead to an improved formula for video popularity and consequently to new findings when analysing popular instructional videos on YouTube.

Directions for Future Research

Over the course of this research, new questions and interesting avenues for future research became apparent. To start with, this study identified in what specific ways instructional design principles can be applied. An investigation using eye-tracking techniques could shed light on the question which of the five types of signaling guides the user's attention most efficiently. Further, the optimal contrast level between the colour of the added visual clues and the user interface could be investigated, which again could lead to improved attention guidance of visual clues. Second, four distinct forms of goal promotion have been discovered and further investigations could identify which ones yields the highest motivational effect on learners. Third, the study indicated that seeing the narrator's face either permanently or in task demonstration free scenes leads to higher video popularity. However, the latter category was only found four times in the sample. It is therefore important to empirical test this finding in a study that employs a bigger sample size. Seeing the narrator's face could also increase the learners' efforts (Fiorella & Mayer, 2018), another effect that could be explored in a follow-up study. Fourth, the analyses of the transcript revealed many flaws in the narratives. Applying the streamlined-step model in the narration design is expected to have positive effects on the users' learning, a hypothesis is worth testing in an empirical study. Also, this research suggests that a new operationalisation of Mayer's segmentation principle is called for as the current operationalisation appears to be unrealistic outside of educational research settings.

References

Alexander, K. P. (2013). The usability of print and online video instructions. *Technical Communication Quarterly, 22*(3), 237–259. https://doi.org/10.1080/10572252.2013.775628.

Anderson, J. R. (1982). Acquisition of cognitive skill. Psychological Review, 89, 369–406.

Bovair, S., & Kieras, D.E. (1991). *Toward a model of acquiring procedures from text*. Handbook of reading research, 2, 206-229.

Brame, C. J., & Perez, K. E. (2016). Effective educational videos: principles and guidelines for maximizing student learning from video content. *Cbe—Life Sciences Education*, *15*(4), 6. https://doi.org/10.1187/cbe.16-03-0125

Brar, J., & van der Meij, H. (2017). Complex software training: Harnessing and optimizing video instruction. *Computers in Human Behavior, 70*, 475–485. https://doi.org/10.1016/j.chb.2017.01.014

Carroll, J. M. (1990). *The Nurnberg funnel. Designing minimalist instruction for practical computer skill*. Cambridge, MA: MIT Press.

Chong, F. (2018). Youtube beauty tutorials as technical communication. *Technical Communication*, *65*(3), 293–308.

Clark, J. M., & Paivio, A. (1991). Dual coding theory and education. *Educational Psychology Review, 3*(3), 149-210.

Clark, R. C., & Mayer, R. E. (2011). E-learning and the science of instruction: proven guidelines for consumers and designers of multimedia learning. San Fransisco, CA: Pfeiffer. http://dx.doi.org/10.1002/9781118255971.

Clement, J. (2019, June 19). Global logged-in YouTube viewers per month 2017-2019. *Statistica*. Retrieved from https://www.statista.com/statistics/859829/logged-in-youtube-viewers-worldwide/

Colman, A. (2008). A Dictionary of Psychology. Oxford University Press. Retrieved 3 Nov. 2020, from https://www.oxfordreference.com/view/10.1093/acref/ 9780199534067.001. 0001/acref-9780199534067.

de Jong, T. (2010). Cognitive load theory, educational research, and instructional design: some food for thought. *Instructional Science*, *38*(2), 105–134.

de Koning, B. B., Tabbers, H. K., Rikers, R. M. J. P., & Paas, F. (2010). Learning by generating vs. receiving instructional explanations:Two approaches to enhance attention cueing in animations. *Computers & Education, 55*, 681-691.

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*. https://doi.org/10.1007/s11423-020-09783-4 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*, (0123456789). https://doi.org/10.1007/s11423-020-09783-4

Ertelt, A. (2007). *On-screen videos as an effective learning tool. The effect of instructional design variants and practice on learning achievements, retention, transfer, and motivation*. (Doctoral dissertation). Albert- Ludwigs Universität Freiburg, Germany.

Farkas, D. K. (1999). The logical and rhetorical construction of procedural discourse*. Technical Communication, 46*, 42-54.

Fiorella, L., & Mayer, R. E. (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

Garrett, N. (2020). Segmentation's failure to improve software video tutorials. *British Journal of Educational Technology*, *(20200719)*. https://doi.org/10.1111/bjet.13000

Guo, P. J., Kim, J., & Rubin, R. (2014, March 4e5). How video production affects student engagement: An empirical study of MOOC videos. In Paper presented at the L@S '14 (Atlanta, GA). https://doi.org/10.1145/2556325.2566239

Haridakis, P.P.D., & Hanson, G.M.A. (2009). Social interaction and co-viewing with youtube: Blending mass communication reception and social connection. *Journal of Broadcasting & Electronic Media,53*(2), 317-335. https://doi.org:10.1080/08838150902908270

Höffler, T. N., & Leutner, D. (2007). Instructional animation versus static pictures: a meta analysis. *Learning and Instruction*, *17*(6), 722–738 . https://doi.org/10.1016/j.learninstruc.2007.09.01

Hong, J., Pi, Z., & Yang, J. (2018). Learning declarative and procedural knowledge via video Lectures: Cognitive load and learning effectiveness. *Innovations in Education and Teaching International*, *55*(1), 74–81. https://doi.org/ 10.1080/14703297.2016.1237371

Johnson, R. B., & Onwuegbuzie, A. J. (2004). Mixed methods research: a research paradigm whose time has come. *Educational Researcher*, *33*(7), 14–26.

Kang, H. J., & Williamson, V. J. (2014). Background music can aid second language learning. Psychol- ogy of Music, 42(5), 728–747.

Kennedy, G. E. (2004). Promoting cognition in multimedia interactivity research*. Journal of Interactive Learning Research, 15*, 43-61.

Khan, M. (2017). Social media engagement: What motivates user participation and consumption on YouTube? *Computers in Human Behavior, 66*, 236-247. doi:10.1016/j.chb.2016.09.024

Laver, J. (1994). *Principles of phonetics*. Cambridge: Cambridge University Press. Lawton, Sherman. leaning performance of learners with different academic backgrounds and ages*. International journal of Scientific Research in Education,13(*2), 283-293.

learning. *Computers & Education, 78*, 150–159. https://doi.org/10.1016/j.compedu.2014.06.003

Lloyd, S.A. & Robertson, C.L. (2012). Screencast tutorials enhance student learning of statistics. *Teaching of Psychology 39*, 67-71.

Mackiwiez, J. (2010). Assertions of expertise in online product reviews. *Journal of Business and Technical Communication, 24*, 3–28. https://doi.org/ 10.1177/1050651909346929

Mayer, R. E. (2003*). Learning and instruction*. Upper Saddle River, NJ: Prentice Hall.

Mayer, R. E. (2005). *The Cambridge handbook of multimedia learning*. Cambridge, NY: Cambridge University Press.

Mayer, R. E. (2014).*Cognitive theory of multimedia learning.* In R. E. Mayer (Ed.), *Cambridge handbooks in psychology. The Cambridge handbook of multimedia learning*n(p. 43–71). Cambridge University Press.https://doi.org/10.1017/CBO9781139547369.005

Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning.*Educational Psychologist, 38*(1), 43- 52. https://doi.org/10.1207/S15326985EP3801_6

Mayer, R. E., Fennell, S., Farmer, L., & Campbell, J. (2004). A Personalization Effect In Multimedia Learning: Students Learn Better When Words Are In Conversational Style Rather Than Formal Style. .*Journal of Educational Psychology, 96*(2), 389–395.https://doi.org/10.1037/0022 0663.96.2.389

Mayer, R. E., Fennell, S., Farmer, L., & Campbell, J. (2004). A Personalization Effect in Multimedia Learning: Students Learn Better When Words Are in Conversational Style Rather Than Formal Style. *Journal of Educational Psychology, 96*(2), 389–395. https://doi.org/10.1037/0022-0663.96.2.389

McHugh, M. L. (2012). Interrater reliability: the kappa statistic. *Biochemia Medica*, *22*(3), 276–282.

Merriënboer, J. J. G., Kirschner, P. A., & Kester, L. (2003). Taking the load off a learner’s mind: Instructional design for complex learning. *Educational Psychologist, 38*(1), 5-13.

Mestre, L. S. (2012). Student preference for tutorial design: A usability study. *Reference Service Review, 40*(2), 258-276.

Miao, M., Zhang, Z., Chen, X., Li, L. (2020). The Effect of Instructional Video Speaking Rate on the Leaning Performance of Learners with Different Academic Backgrounds and Ages, *13*(3), 283–293.

Morain, M., & Swarts, J. (2012). Yoututorial: a framework for assessing instructional online video. *Technical Communication Quarterly*, *21*(1), 6–24. https://doi.org/10.1080/10572252.2012.626690

Neelen, M., & Kirschner, P. A. (2020).*Evidence-informed learning design: creating training to improve performance*. Kogan Page, Limited.

O'Neil-Hart, C. (2017, December). *Why you should lean to how-tocontet in 2018*. Think with Google. https://www.thinkwithgoogle.com/marketing-strategies/video/self-directedlearning-youtube/

Paivio, A. (1986). Mental representations: *A dual coding approach*. Oxford, UK: Oxford University Press.

Palmiter, S. L., Elkerton, J., & Baggett, P. (1991). Animated demonstrations vs written instructions for learning procedural tasks: A preliminary investigation. *International Journal of Man-Machine Studies, 34*(5), 687–701. https://doi.org/10.1016/0020-7373(91)90019-4

Palmiter, S., & Elkerton, J. (1993). Animated demonstrations for learning procedural computer-based tasks. Human-Computer Interaction, 8(3), 193–216. https://doi.org/10.1207/s15327051hci0803_1

Payne, S. J., Chesworth, L., & Hill, E. (1992). Animated demonstrations for exploratory learners. *Interacting with Computers, 4*, 3–22.

Payne, S. J., Chesworth, L., & Hill, E. (1992). Animated demonstrations for exploratory learners. *Interacting with Computers, 4*, 3–22.

Richards, D., Fassbender, E., Bilgin, A., & William, F. T. (2008). An investigation of the role of back-ground music in IVWs for learning. Research on Learning Technology, 16(3), 231–244. https ://doi. org/10.3402/rlt.v16i3 .10901 .

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. J. Educ. Psychol. 76, 647–658. https ://doi. org/ 10.1037/0022-0663.76. 4.647

Schunk, D. H. (1996*). Learning theories* (2nd ed.). Englewood Cliffs, NJ: Prentice-Hall.

Schwan, S., & Riempp, R. (2004). The cognitive benefit of interactive videos: Learning to tie nautical knots. *Learning and Instruction, 14*, 293-305.

Shoufan, A., & Mohamed, F. (2017). On the Likes and Dislikes of YouTube's Educational Videos, (4), 127 132. https://doi.org/10.1145/3125659.3125692

Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research, 78*(1), 153-189.

Snelson, C., Rice, K. & Wyzard, C. (2012). Research priorities for YouTube and video-sharing technologies: A delphi study. British Journal of Educational Technology, *43(*1). https://doi.org/10.1111/j.1467-8535.2010.01168.x

Spannagel, C., Girwidz, R., Löthe, H., Zendler, A., & Schroeder, U. (2008). Animated demonstrations and training wheels interfaces in a complex learning environment. *Interacting With Computers, 20*, 97–111.

Swarts, J. (2012). New modes of help: Best practices for instructional video. *Technical Communication, 59*(3), 195-206.

Sweller, J. (1988). Cognitive load during problem solving: effects on learning. *Cognitive Science*, *12*(2), 257–285. https://doi.org/10.1207/s15516709cog1202_4

Sweller, J. (1989). Cognitive technology: some procedures for facilitating learning and problem solving in mathematics and science. *Journal of Educational Psychology*, *81*(4), 457–466. https://doi.org/10.1037//0022-0663.81.4.457

Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction*, *4*(4), 295–312. https://doi.org/10.1016/0959-4752(94)90003-5

Sweller, J., Ayres, P. L., & Kalyuga, S. (2011). *Cognitive load theory* (Ser. Explorations in the learning sciences, instructional systems and performance technologies). Springer. https://doi.org/10.1007/978-1-4419-8126-4

Sweller, J., van Merriënboer Jeroen J. G, & Paas, F. (2019). Cognitive architecture and instructional design: 20 years later. *Educational Psychology Review*, *31*(2), 261–292.

Tauroza, S., & Allison, D. (1990). Speech rates in British English. *Applied Linguistics, 1*(1). 90 – 105. https://doi.org/10.1093/applin/11.1.90

Ten Hove, P., & van der Meij, H. (2015).Like it or not. What characterizes youtube's more popular instructional videos? *Technical communication*, *62*(1), 48-62.

Think with Google (2017). *How-to videos earn the most attention of any content category on YouTube, even more than music clips or gaming.* https://www.thinkwithgoogle.com/consumer-insights/consumer-trends/how-to-videos-attention-statistics/

Tversky, B., Bauer-Morrison, J., & Betrancourt, M. (2002). Animation: Can it facilitate? *International Journal of Human-Computer Studies, 57*(4), 247-262. https://doi.org/10.1006/ijhc.2002.1017

Van der Meij, H., & Carroll, J. M. (1998). Principles and heuristics for designing minimalist instruction. In J. M. Carroll (Ed.), *Minimalism beyond the Nurnberg funnel* Cambridge, Mass: MIT Press.

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.

van der Meij, H., & van der Meij, J. (2014). A comparison of paper-based and video tutorials for software learning. *Computers & Education, 78*, 150–159. https://doi.org/10.1016/j.compe du.2014.06.003

van der Meij, J., & van der Meij, H. (2015). A test of the design of a video tutorial for software training. *Journal of Computer Assisted Learning*, *31*(2), 116–132.

van Wermeskerken, M., Ravensbergen, S., & van Gog, T. (2018). Effects of instructor presence in video modeling examples on attention and learning. *Computers in Human Behavior, 89,* 430–438. https://doi.org/10.1016/j.chb.2017.11.038

Walker, & Gareth.(2010). The phonetic constitution of a turn-holding practice: Rush-throughs in Eeglish talk-in-interaction. In Dagmar Barth-Weingarten, Elisabeth Reber & Margret Selting (eds*.),* *Prosody in interaction*, 51–72. Amsterdam & Philadelphia: John Benjamins. https://doi.org/10.1075/sidag.23.08wal

Worlitz, J., Stabler, A., Peplowsky, S., & Woll, R. (2016). Video tutorials: An appropriate way of teaching Quality management tools applied with software. *Quality Innovation Prosperity, 20*(2), 169-184. https://doi.org/10.12776/ QIP.V20I2.754.

Wouters, P., Tabbers, H. K., & Paas, F. (2007). *Interactivity in video-based models. Educational Psychology Review, 19*, 327-342.