Computers in Human Behavior 70 (2017) 475–485

Contents lists available at ScienceDirect

Computers in Human Behavior

journal homepage: www.elsevier.com/locate/comphumbeh

Full length article

Complex software training: Harnessing and optimizing video instruction

Jagvir Brar [a], Hans van der Meij [b, *]

[a] University of Twente, Faculty of Behavioural, Management and Social Sciences, Educational Science and Technology, P.O. Box 217, 7500, AE Enschede, The Netherlands
[b] University of Twente, Faculty of Behavioural, Management and Social Sciences, Department of Instructional Technology, P.O. Box 217, 7500, AE Enschede, The Netherlands

ARTICLE INFO

Article history:
Received 21 October 2016
Received in revised form
6 January 2017
Accepted 7 January 2017
Available online 9 January 2017

Keywords:
Video
Statistics
Demonstration-Based Training (DBT)
Reviews

ABSTRACT

This article investigates the design and effect of optimized video for statistics instruction. In addition, the use of video reviews to further optimize video instruction is examined. A Demonstration-Based Training (DBT) model was proposed and followed for the construction of the video. The videos were tested in a university-level statistics course. Students were randomly assigned to an experimental condition with demonstration and review videos and control condition with only demonstration videos. Video activity was logged to collect engagement data (coverage and commitment), and a knowledge and performance test were administered. The data showed that the videos were successful at gaining and maintaining the motivation and attention of students. Knowledge scores were moderate and there was no main effect for condition. Regression analysis showed overall coverage and review commitment were predictors for knowledge scores. Performance scores remained high when compared to the previous cohort, however there was a significant positive difference in the current study. There was no main effect for condition on performance scores. The DBT-model and its implementation in the videos was considered successful. In addition, it is suggested that video instruction can play an important role in statistics courses where theory and practice are separated.

1. Introduction

Statistics is widely taught in higher education. Understanding it involves various theories, concepts, and formulas. Statistics software has added another layer of required understanding from students. Statistics software is a useful tool for automating calculations and creating graphical displays, and its use for statistical analysis is the new norm (Baglin & Da Costa, 2014). There is a wide range of statistics software available (e.g., SAS, Minitab, R, STATA, SPSS), and students of modern statistics courses are required to develop some level of technological skill alongside statistical understanding.

Appropriate instruction, such as video, can be used to help students improve understanding of statistics. Videos are already being harnessed for instruction (Lloyd & Robertson, 2012), especially in higher education (Kay & Kletskin, 2012). Various studies have also found that video instruction is effective to support learning outcomes (Kay & Kletskin, 2012; Lloyd & Robertson, 2012; van der Meij, 2014), and students have described it as being enjoyable to watch, motivating, and helpful (Kay, 2014).

Video instruction that allows viewers to learn through observing a behavior model may be especially beneficial for students learning statistics and statistics software. This can take the form of Demonstration-Based Training (DBT), where instructional features complement a model of task performance. Ideally this would result in learners understanding the software while grasping statistics concepts, allowing them to reproduce the behavior for future tasks.

Thus, the present article aims to design an optimized video to help students in higher education learn statistics and learn how to use statistics software. As in many other universities, the statistics courses in our university generally start with lectures that focus on statistics facts and concepts. These lectures are then followed by practicals that focus on task performances that require students to use statistics software to solve problems. The practicals rely on, but do not cover, statistical theory, which causes difficulties for many

* Corresponding author.
E-mail address: H.vanderMeij@utwente.nl (H. van der Meij).

http://dx.doi.org/10.1016/j.chb.2017.01.014

students. The video instruction that we designed set out to improve the linkage between theory and practice. The instruction should convey statistics foundations and their application, along with software usage. It was tested for learning and motivation within a real setting.

2. A demonstration-based training (DBT) approach to software training

For constructing a video tutorial on software training, we adopted a Demonstration-Based Training approach. DBT originates in Bandura's (1986) views on observational learning. Per DBT, observational learning depends on a model of performance that is complemented with instructional features. Recently, a theoretical model of DBT has been advanced for trainer-led management training (Grossman, Salas, Pavlas, & Rosen, 2013; Rosen et al., 2010). This paper proposes an adapted model that fits the aims and context of software training (see Fig. 1).

The main tenet of the DBT model is that the instructional features in design should address the four interrelated processes of: attention, retention, production, and motivation. To create a baseline level of support for the user, it is necessary to implement at least one design guideline to support each process. The model shows the wide range of design options that can be followed to achieve this.

The theoretical model further considers user characteristics and situational variables. User characteristics can influence the design and effectiveness of a video. Multimedia research shows that especially prior knowledge is an important characteristic (Mayer, 2014a). Users with low prior knowledge often benefit considerably from instructional support that is added to a model of performance. In contrast, users with high prior knowledge often do not need, and may even be hampered by such scaffolding (Kalyuga, 2007). Situational variables have implications for the specific design guidelines that can be applied. For instructional video on software training, we assume that the prevalent context is a solitary user working on his or her computer. This means that design features that involve instructor-led support, among others, are not an option. The next section describes the four processes in observational learning and the design guidelines for supporting each process.

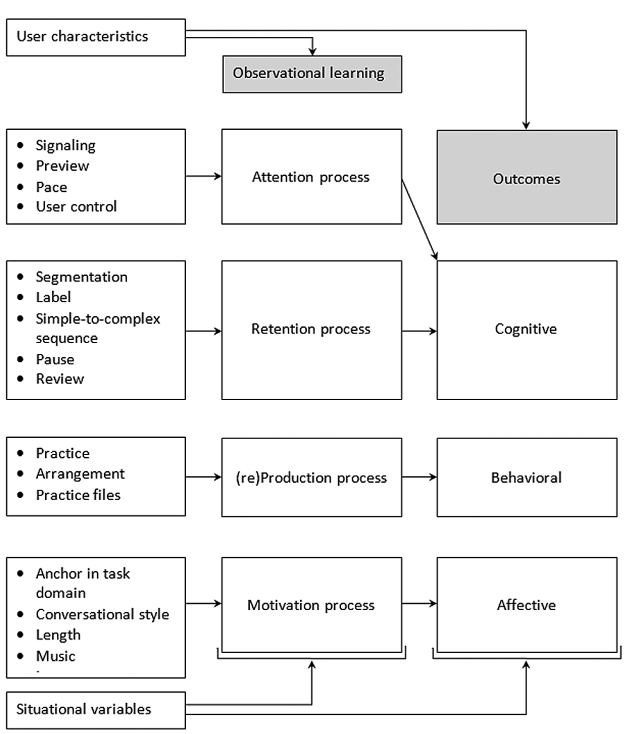

Fig. 1. DBT-model of the connection between conditions, instructional features, learning processes and outcomes in software training.

2.1. Attention

Attention is an active process in which the user filters and selects information (Anderson, 2010). Information that does not pass this initial phase of processing is lost. To learn from a demonstration, the user must therefore attend to and accurately perceive the modeled performance.

Attention is influenced by the aspects of *space* and *time* (Smith & Kosslyn, 2007). Space refers to the amount of information that the user can process simultaneously. Time refers to the speed in which new information is presented to the user. Both aspects influence attentional processes in software training. Regarding space, an important factor is the complexity of the interface. A new interface can pose a considerable challenge to the user's attentional processes (Shneiderman, Plaisant, Cohen, & Jacobs, 2010). Regarding time, the transience of video is often mentioned as a possible barrier (Brucker, Scheiter, & Gerjets, 2014; Lowe, Schnotz, & Rasch, 2011). The fleetingness of video can tax the users' attentional processes; new information may follow before the user has been capable of processing what has just been presented.

Attention can be a *bottom-up* or *top-down* process (Anderson, 2010). In the latter case, attention is guided by what the user already knows of the topic. That can be the existence of prior knowledge or newly acquired knowledge from a preview. Attention is a bottom-up process when it is stimulated by a physical feature of the video (e.g., visual or auditory signal).

2.1.1. Attention guidelines

A well-known means of supporting attentional processes in a bottom-up fashion is cueing or ***signaling*** (A1). Signals attend the user to key points of information in a presentation without adding content (Lemarié, Lorch, Eyrolle, & Virbel, 2008). Signals such as a color coding or the presence of an arrow-shaped or circled overlay can raise the salience of a location or object. Eye movement records of users give empirical evidence of the attention-directing effect of signals (e.g., Boucheix & Lowe, 2010; Kriz & Hegarty, 2007; de Koning, Tabbers, Rikers, & Paas, 2010). A recent meta-analysis concluded that signals in multimedia presentations significantly raise learning (Richter, Scheiter, & Eitel, 2015).

Where signaling mainly affects attention in a bottom-up fashion, the presence of a ***preview*** (A2) aims to support attentional processes top-down. A preview is a short presentation before a demonstration that informs the user about the goal, jargon, and/or identifies and locates important objects. A preview-related design guideline from multimedia research is the pre-training principle. This principle holds that multimedia learning improves when people have received short advanced instructions about the names and characteristics of the main concepts in an instruction. There is considerable empirical support for the pre-training principle (e.g., Mayer & Pilegard, 2014; Mayer, 2011).

An elusive and yet important aspect regarding the user's attentional processes is the speed or ***pace*** (A3). Pace is the ongoing stream of information in a video. It is determined by the speed of presentation for content, and formal elements such as shots, transitions, graphics, and audio. Individually and interactively these aspects affect cognitive and motivational processes of the user. Accordingly, empirical research reports measures of pace that concern design features (e.g., fast versus slow cuts), presentation speed (slow or fast plays of clips), and system versus learner-pace (e.g., Bracken, Pettey, Guha, & Rubenking, 2010; Meyer, Rasch, & Schnotz, 2010; Stiller, Freitag, Zinnbauer, & Freitag, 2009). Both Koumi (2013) and van der Meij and van der Meij (2013) recommended opting for a moderate pace. A pace that is too fast causes cognitive overload (Lang, Park, Sanders-Jackson, Wilson, & Wang, 2007). A pace that is too slow makes a video boring which can decrease attention. Empirical research on pace has yielded equivocal findings (e.g., Boucheix & Guignard, 2005; Boucheix, Lowe, & Bugaiska, 2015; Simonds, Meyer, Quinlan, & Hunt, 2006). This is presumably caused by the fact that pace often significantly interacts with other factors such as user characteristics (Berney & Bétrancourt, 2016).

Most video players nowadays include a toolbar that gives ***user control*** (A4). Options such as a stop, pause, and rewind allow users to influence the pace of the video and make it better fit their capacity for taking in information. This is illustrated by Schwan and Riempp (2004) who reported a heftier use of the interactive features of the toolbar for more difficult tasks. In addition, user control is valued because it gives users the opportunity for selectively viewing a video based on their learning needs (Schreiber, Fukuta, & Gordon, 2010). Several empirical studies have reported positive effects of user control on learning from dynamic visualizations (e.g., Hasler, Kersten, & Sweller, 2007; Höffler & Schwartz, 2011; Merkt & Schwan, 2014; Merkt, Weigand, Heier, & Schwan, 2011; Stiller et al., 2009; Witteman & Segers, 2010). However, a recent meta-analysis on animations could not validate this finding and indicates that more research is needed to discover how user control affects learning processes (Berney & Bétrancourt, 2016).

2.2. Retention

Retention revolves around the processes of understanding and storing information for future behavior. In retention, the user must transform incoming information into symbolic codes and store these in long-term memory (Bandura, 1986). Mayer (2014b) refers to these processes respectively as organization and integration. To build understanding, the user must organize the selected information into a coherent cognitive structure in working memory. For procedures, such a structure may have the form of a narrative because experience is usually organized as such (Wagoner, 2008; Zacks, Speer, Swallow, Braver, & Reynolds, 2007). The effect of the comprehension process is a succinct, prototypical representation of task performance. The outcome must be committed to memory to serve as a guide for subsequent action. This requires that the user integrates new with old information. The newly formed cognitive structure must be connected to existing prior knowledge in long-term memory. The result of the retention process is a concept of task performance that can serve as a guide and standard, enabling the user to organize, initiate, and monitor future actions.

2.2.1. Retention guidelines

An instructional feature that can support the retention process is ***segmentation*** (R1). Segmentation involves dividing a longer video into several clips or within-video sections containing a clear beginning and end. Segmentation is preferably based on a meaningful division of tasks or concepts. Empirical studies have repeatedly found that segmentation enhances learning from multimedia (e.g., Catrambone, 1995, 1998; Margulieux, Guzdial, & Catrambone, 2012; Schittek Janda et al., 2005; Schwan, Garsoffky, & Hesse, 2000).

An instructional feature that serves as a visible sign of segmentation within a clip is a ***label*** (R2). A label summarizes the key point of a short video section. The presence of labels in multimedia creates a 'desirable difficulty' with the spoken narrative. Research has found this instructional feature conducive for learning from multimedia presentations. By creating a small discrepancy between what the narrator says and what is written on the screen, the user is stimulated to pay close attention to both sources and process the information at a deeper level (Bjork & Bjork, 2011; Yue, Bjork, & Bjork, 2013). In addition, all labels together give a structural overview of the content of a clip which should promote retention.

Segmentation may lead to the problem of sequencing. Even when there is an obvious, general dictate from the tasks or concepts that are taught, sequencing of specific clips can be problematic. The design guideline to follow in these cases is adoption of a ***simple-to-complex sequence*** (R3). In such a sequence, easier tasks are presented early on and more complex tasks appear later. Empirical research shows that a simple-to-complex sequence is beneficial for learning (e.g., Clarke, Ayres, & Sweller, 2005; Pollock, Chandler, & Sweller, 2002). The study from Clarke et al. is particularly relevant for the present study because it showed that it is better to adopt a sequential, rather than mixed, presentation when users must learn both about software (i.e., a spreadsheet program) and a domain (i.e., mathematics).

Another instructional feature that can contribute to retention is the inclusion of ***pauses*** (R5). Pauses are short, two to 5 s breaks within a video. During a pause, no new (visual or auditory) information is presented, which gives the user time to digest what has already been presented. In addition, pauses demarcate concept or event boundaries which helps in organizing the information. Empirical research shows that pauses yield better (i.e., lower) difficulty ratings for dynamic representations and significantly raise learning (e.g., Hassanabadi, Robatjazi, & Savoji, 2011; Lusk et al., 2009; Spanjers, van Gog, Wouters, & van Merriënboer, 2012; Spanjers, Wouters, van Gog, & van Merriënboer, 2011).

An instructional feature that can also support retention is the ***review*** (R6). A review is a concise recap of the main information. Reviews have hardly been investigated in multimedia research. Because the present study manipulates their presence to create two video conditions, we describe reviews in a separate section that follows the discussion of the general DBT-framework.

2.3. Production

The goal of observing a demonstration is that the user can solve a problem or accomplish a task. To (re)produce what has been learned, the user must be able to recall or reconstruct the solution steps and monitor their correct execution (Bandura, 1986).

2.3.1. Production guidelines

Facilitating the user's production processes largely falls outside the realm of video construction. What designers can do however, is to stimulate ***practice*** (P1). Practice is a user action that benefits learning in two ways (Leppink, Paas, van Gog, van der Vleuten, & van Merriënboer, 2014; van Gog, Kester, & Paas, 2011). It can consolidate retention and reinforce what the user remembers. In addition, it can bring errors or omissions to attention. During practice, the user may discover that a mistake is made or something is forgotten. This awareness may prompt the user to restudy (sections of) the video.

Research on the optimal ***arrangement*** (P2) for practice indicates that there is often a trade-off between immediate success and learning. That is, when practice follows immediately after instruction, the success rate is higher than when there is massed practice at the end. In other words, when job-aiding is the primary goal, it is best to opt for immediate practice after a video clip. When the goal is learning and transfer, massed practice after a series of clips is better (e.g., Helsdingen, van Gog, & van Merriënboer, 2011; Schmidt

& Bjork, 1992).

The user's production process efforts can be facilitated by making ***practice files*** (P3) available. Such files enable the user to focus on what must be learned by representing a prototypical problem or task and keeping distracting information to a minimum (van der Meij & Carroll, 1998).

2.4. Motivation

Motivation is the driving force behind the processes of attention, retention, and production. It can be defined as the process where goal-directed activity is instigated and sustained (Pintrich & Schunk, 2002). Bandura (1997) claims that task value appraisals and self-efficacy judgments play a mediating motivational role in learning from a model. In the CANE model of motivation, these two aspects fall under the broader rubrics of mental effort and commitment (Clark, 2015). Factors that stimulate users to spend mental effort concern effectiveness values such as utility, interest, and importance. Factors that stimulate users to commit themselves and engage in active and sustained goal pursuits are self-efficacy, mood, and personal value.

2.4.1. Motivation guidelines

An instructional feature that can support motivation in software training is ***anchoring the tool in the task domain*** (M1) (van der Meij & Carroll, 1998). For most users, a software program is a tool they want to use to achieve objectives in their task domain. The program is merely a means; it is almost never an end. To the greatest extent possible, tasks should therefore be selected from the core tasks of the user's application domain. Domain anchoring is one of the four main principles in the minimalist approach to software training. Empirical research has repeatedly established the effectiveness of this approach for user motivation (e.g., self-efficacy) and procedural knowledge development in software training (Carroll, 1990, 1998).

The narrative in a video can enhance user motivation by adopting a ***conversational style*** (M2). This style is characterized by the frequent usage of personal pronouns (e.g., I, you, we) and the occasional presence of self-revealing comments of the designer. In this way, a sense of social partnership is created between the designer and the user. The design guideline is best known under the term personalization principle. A recent meta-study found that a conversational style yields more learning than a formal style, but this effect disappeared for instructions of longer than 35 min (Ginns, Martin, & Marsh, 2013). The meta-study further concluded that conversational style had only a small effect on motivation. However, a more recent empirical study achieved a significant motivational effect with such a style (Reichelt, Kämmerer, Niegemann, & Zander, 2014).

Several large-scale studies show that long videos may lead to premature dropout (Guo, Kim, & Rubin, 2014; Wistia, 2012). To prevent users from navigating away from a video before completion, it is probably best to aim for a **length** (M3) of 3–5-min maximum.

Practitioners have long since advocated the inclusion of background ***music*** (M4) for its positive effects on motivation in multimedia designs. For instance, the inclusion of music is strongly advocated for achieving a mood setting effect in movie trailers (Finsterwalder, Kuppelwieser, & de Villiers, 2012). A recent empirical study also found a significant effect of the type of trailer music on the viewer's feelings (Strobin, Hunt, Spencer, & Hunt, 2015). In addition, the authors suggested that the effect was subconscious, persuading viewers to feel in a certain way. Viewers seemed unaware of the motivating influence of the music. Likewise, Koumi (2013), a long-term consultant on audio and video for the UK Open University, claims that background music can enhance motivation. In addition, he adds that it requires a considerable amount of time to discover the right kind of music. It should fit the mood, but not interfere with the story. Finally, a recent analysis of instructional videos on YouTube revealed that the presence of background music distinguished popular videos from average videos (ten Hove & van der Meij, 2015). Empirical research on background music in instructional multimedia designs is scarce. A meta-study of the few experiments that compared conditions with and without music concluded that music had a small, detrimental effect on cognition, but a positive effect on emotion (Kämpfe, Sedlmeier, & Renkewitz, 2010).

3. Video reviews

An instructional feature that can also support learning but has received little attention is the review. The lack of attention is odd as there are good conceptual reasons for including a review in an instruction. An important argument in favor of reviews is that they can support retention by summarizing the key points for the user. Especially for viewers with low summarization skills, this can be very helpful. In addition, the review can serve the user as a frame of reference for his or her own summary. The predefined and self-constructed review can be compared. A third argument is that a review is a kind of rehearsal that can strengthen the memory trace.

Also, empirical evidence favoring the inclusion of reviews is accumulating. There is both indirect and direct empirical support for the contention that reviews benefit learning. Indirect support comes from three recent experiments on videos (without reviews) for software training. These studies revealed that it is difficult to achieve mastery levels for learning (van der Meij & van der Meij, 2014; van der Meij & van der Meij, 2015; van der Meij, 2014). In these studies, the findings during training attested to the basic quality of the videos. That is, on tasks attempted during training when users could still consult the videos, success rates of well over 80 percent were achieved. However, after training when users could no longer consult the videos, the average task achievement success dropped to below 70 percent. In short, these studies suggest that demonstration videos that effectively enhance aided task performances are likely to need special design measures to enhance retention and learning. The inclusion of a review could be one such measure.

Direct empirical support comes from two recent experiments on video reviews in software training (van der Meij & van der Meij, 2016a,b). In these studies, an experimental condition with reviews was compared to a control condition without reviews. Both studies found significant and substantial effects of reviews on learning. In addition, reviews positively enhanced motivation. The videos in these studies revolved around Microsoft Word's formatting options, supporting primarily procedural knowledge development. The videos in the present study differ in content and aim. They revolve around statistics and seek to support both conceptual and procedural knowledge development.

4. Research design and questions

The empirical study included an experimental condition with demonstration and review videos and a control condition with only demonstration videos. The real setting in which this study was conducted limited the research questions for which data could be gathered. These questions were the following:

Question 1: How engaging are the videos? The motivating qualities of the videos were assessed by looking at two measures of engagement, namely coverage and commitment. Coverage refers to the proportion of the videos that may have been viewed. It was operationalized as the percentage of the total amount of video seconds played at least once. Commitment refers to the duration that the videos have been played. It is expressed as a percentage of the total length of the videos.

Question 2: How effectively do the videos support knowledge development? A knowledge test assessed the students' declarative knowledge of statistics. The primary objective of the videos was to enable all students to achieve a pass score. In addition, the assumption was tested that the experimental condition yielded a significantly higher knowledge test score than the control condition. Exploratory regression analyses were performed to find explanatory factors for the knowledge test scores.

Question 3: How well do the videos support performances? The practical is a performance test in which the students must integrate theoretical knowledge of statistics with knowledge on using SPSS software. The study investigates whether all students achieved a pass score on their performance test. The performance test in the study was the same as in the previous cohort. That score was therefore used as a benchmark for evaluating whether the performance of all students in the experiment was higher. In addition, the assumption was tested that the experimental condition yielded a significantly higher performance test score than the control condition. Exploratory regression analyses were performed to find explanatory factors for the performance test scores.

5. Method

5.1. Participants

The target audience consisted of pre-Master students enrolled in an obligatory Inferential Statistics course (n = 133) at a University in the Netherlands. Participation in the experiment was voluntary and students could opt-out at any time during the study. No course credits could be won for participation. Students were randomly but evenly assigned to conditions. Only students that logged on to watch the videos and completed the knowledge test were included in the data analysis. The participation rate was 82% with 110 participants (61 female, 49 male), aged 20–46 (M = 24). 75% of the participants were of Dutch nationality, 12% were German, and the remaining 13% were of other nationalities. One student was removed due to technical issues, resulting in a final data set of 109 participants (54 in the experimental condition, 55 in the control condition).

Data from the previous cohort was also used, which had 178 students (93 female, 85 male), aged 19–58 (M = 23). Most students were again of Dutch nationality (73%), 18% were German, and the remaining 9% were of other nationalities. It was assumed that the previous students did not differ much in terms of motivation.

5.2. Instructional materials

The video clips explained what a *t*-test is and how to conduct one using IBM SPSS statistics software (version 23). The construction of the clips was based on extensive content and task analyses. These analyses included Andy Field's (2013) chapter on t-tests from "Discovering Statistics Using IBM SPSS statistics". This textbook won the British Psychological Society Book Award in 2007 and was, for a long time, ranked the number one best-seller on Amazon for books on Mathematical and Statistical Software.

Two main considerations determined the basis of the model that formed the backbone of the videos. One, a decision was made concerning the *solution method* that should be demonstrated. Software programs often provide multiple ways to achieve a task. It is recommended to present only one of these method, and to select the easiest one (Carroll & van der Meij, 1998; Renkl, 2014). For the SPSS videos, this meant, among others, that users were taught how to use the menu for creating the syntax of a statistical procedure rather than writing this syntax themselves. Two, the second major design decision concerned the *discussion of content*. Procedural discourse about software generally consists of information on goals, interactions, prerequisites, and troubleshooting (Farkas, 1999; van der Meij & Gellevij, 2004; van der Meij, Blijleven, & Jansen, 2003). The first two are always needed in user instructions and this information also formed the major part of what was communicated. For the SPSS videos, goal information identified the performance task for the user and occasionally convinced the user of the value of task achievement. Information about interactions concerned the user's actions and ensued software reactions. The action component informed the user on what to do. A preferred format is a succinct statement built around an imperative verb (Farkas, 1999). The reaction component should help the user in perceiving and understanding the effect of an action (van der Meij & Gellevij, 2004).

The complementary design guidelines advanced in the theoretical model shown in Fig. 1, have been followed in the construction of the videos, whenever possible. Their implementation is described below, together with the name and referent of the guideline. Unless stated otherwise, these design guidelines were implemented in all video clips.

The content for the videos was considered too complex for treatment in a single video. This prompted a *segmentation* (R1) that was based on a logical division of the three main events entailed in performing a *t*-test, namely: (1) understanding the basic facts and concepts about t-tests, (2) conducting a *t*-test in SPSS, and (3) understanding and evaluating the SPSS findings.

The title for the videos, "My spouse is a goat", described the overall theme. It abbreviated the main research question being tested, namely whether or not people who are married to a goat are more intelligent than people not married to a goat. Just as in Andy Field's textbook, testing thus involved an odd, imaginary topic. The 'hook' should capture and sustain user interest (see Koumi, 2013).

Clip #1 was subtitled "Introducing the *t*-test". The clip was an exposition of statistical facts and concepts. It reminded students of key facts, and served to build or enhance understanding of what a *t*-test entails. The information in this clip was presented in a *simple-to-complex sequence* (R3). That is, the clip began by discussing the simple problem of a comparison of the means of two independent groups, and then progressively went into a discussion of more complex topics such as sampling, distributional changes, and components in the formula for a *t*-test.

The clip featured 2-s *pauses* (R5) during topical transitions. Fig. 2 illustrates such a transition. The screen in the middle represents the pause. Fig. 2 also shows how *labelling* (R2) was applied. Labels appeared in a small, yellow oval-shaped box on the left bottom of the screen. The box presented labels that consisted of maximally two written words (e.g., equation, hypothesis, population, sample group, T-distribution). During pauses, no label was presented.

The first clip was expected to yield a pre-training effect (*preview*, A2) for the second and third video. That is, it built prerequisite knowledge for the task performance with SPSS and for interpreting the SPSS output. Clip #1 had a duration of 2 min 59 s (plus 37 s for the review).

Clip #2 was subtitled "Performing a *t*-test in SPSS". The clip revolved around a demonstration of the task of achieving a *t*-test using SPSS. Completing a statistical test in SPSS can pose a considerable challenge because the procedure regularly calls upon the user's conceptual knowledge for making the correct menu-choices from the interface. For instance, at one moment during task execution, the user was called upon to recall what a tested or dependent variable entails, to apply that knowledge for the specific

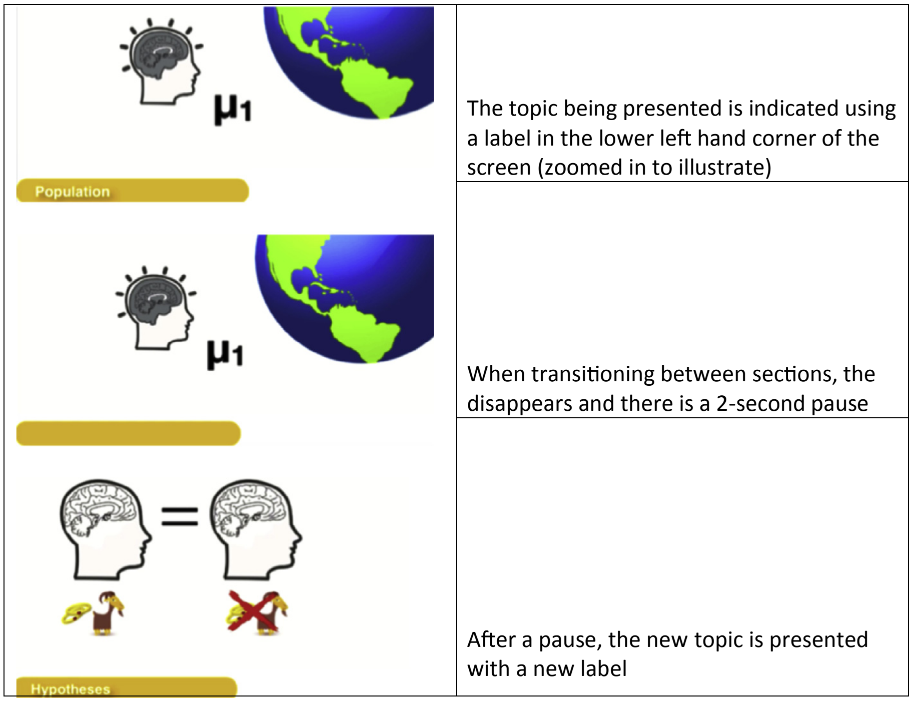

Fig. 2. An illustration of the use of labels and pauses within a video clip.

study, and to make the proper selection of menu-based choices on the interface.

Both the demonstration and the narrative supported the user in making the right moves. The spoken voice in the narrative was that of a male native speaker who used a *conversational style* (M2). That is, the speaker regularly addressed the user as "you" (e.g., "So you've collected the data …", "You're done", and "What you want to know"), regularly made his presence known (e.g., "I won't go into detail", and "I wanted to compare …"), and occasionally vented his opinion (e.g., "… and believe me it's not that complicated", and "Now for the fun part").

Signaling (A1) supported attentional processes (see Fig. 3). Zooming focused on the active part of the interface (i.e., the Independent-Samples T-test dialog box). Signals were drawn on top of the interface, which gave these a dynamism that attracted attention. In addition, a warm color (i.e., red) was used to enhance

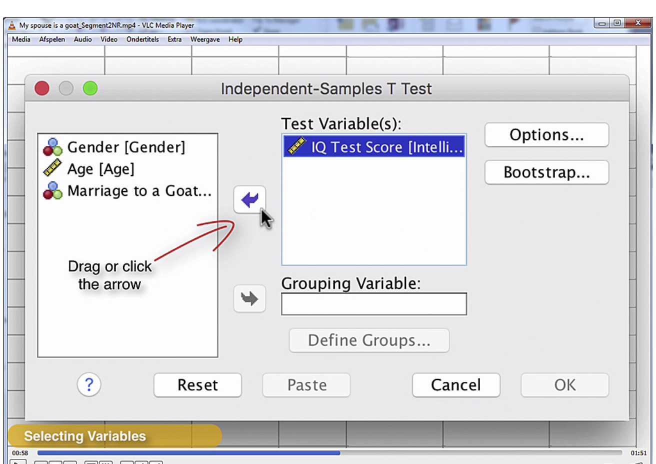

Fig. 3. An illustration of the signaling used in a video clip.

the attention drawing effect (see Kosslyn, Kievit, Russell, & Shephard, 2012). Clip #2 had a duration of 1 min 51 s (plus 29 s for the review).

Clip #3 was subtitled "Understanding SPSS output from a *t*-test". The clip instructed users on where they could find the main results from the *t*-test in the output report that SPSS delivered. In doing so, there was an explanation, which was sometimes a brief repetition of key concepts involved in meaning making (e.g., p-values, Levene test).

The pace (A3) in all three videos largely depended on the speed of the narrative. To obtain a precise measure of pace, the mean words-per-minute (wpm) was measured. The wpm-metric was computed by dividing the total number of narrated words by the total time of a video (in seconds), and then the outcome was multiplied with 60. The wpm count for the three video clips yielded a finding of respectively 145, 133, and 142 wpm. Since the average rate of human speech is 125–150 wpm (Fulford, 1992), the pace of the videos can be qualified as moderate. This is an appropriate speed considering the outcome of a controlled study on pace (Simonds et al., 2006). In that investigation, a videotaped lecturer presented information in one of three different speaking rates (i.e., 116, 172, and 213 wpm). The speaking rate did not affect cognition, but a significant effect on motivation was found. The slow condition yielded lower appraisals than the moderate or fast paced lesson which did not differ.

Music (M4) was presented only during the start and closing of each clip. At the start, the music accompanied a spoken word of welcome and a one-sentence description of the goal (e.g., "Welcome back …. Now you are going to make sense of your SPSS output"). The music (drawn from Creative Commons) played a tune in an upbeat jazzy style for 7–9 s. The same entry music was played in all three video clips. At the close, a tune was played in an instrumental, uplifting style. This tune too was the same in all video clips. The closing music accompanied a spoken wrap up and, for the first two clips, a short preview of the next (e.g., Clip #2: "You're done! You've just ran a *t*-test. In the next segment, you'll learn how to decipher the output SPSS gave you from running this test."). The closing took between 5 and 10 s, and the closing music remained in the background during the reviews.

Clip #3 had a duration of 2 min 56 s (plus 35 s for the review). Just like the other clips, the *length* (M3) of the videos without reviews was below the 3-min threshold suggested by a large-scale study of lectures (Guo et al., 2014). The mean length of the three clips without reviews was slightly above the 2 min 30 s criterion that the Motion Picture Association of America (MPAA) sets for movie trailers (Cross, 2010). Trailers must compete in an open market for consumer attention, where they must subsequently lure the audience into viewing the whole film or documentary. With instructional video, the competition for attention is less fierce, but even so it seems wise to adopt this strict standard for length in video production.

Reviews (R6) presented the most important ideas from a video clip. Their length in clips 1, 2 and 3 was respectively 11%, 10%, and 10% of the demonstration. This seems about the right amount of reduction required for a summary. Each review was structured as a series of steps because this is essentially how each clip had been framed. Reviews appeared after closure of the demonstration. Each review was introduced by the standard phrase "But first, let's review the key points", after which the summary followed (e.g., Clip #2: "To do an independent samples *t*-test in SPSS, I navigated to the main dialog box … I selected what I wanted to compare … And I selected and defined how my sample groups differ … After running the test, SPSS gave me my output"). Reviews can include stills as well as animations. Each review was presented on a yellow background to make it stand out.

The videos and the knowledge test were presented on a website. Students could gain access to this website with their student email address. The website gave students the freedom to choose the videos to play, and whether and when to take the knowledge test. After selecting a video, a still with the title and subtitle of the clip appeared. Students could then start the video by selecting the play button from a standard toolbar. This toolbar also enabled other types of *user control* (A4) such as (temporary) stops and rewinding.

5.3. Instruments

User logs. The students' actions on the videos were unobtrusively recorded with Sprout Video, a video-hosting platform that tracks the actions of every individual user. The most important data from the platform concern the status of each second of a video (e.g., play, replay, and skip). With these data, it was possible to assess video engagement (i.e., coverage and commitment).

Coverage was the number of video seconds that were presented at least once in play mode. It was expressed as a percentage of the total number of seconds in a video. For example, if a student watched the first 100 and last 10 s of Clip #1 (179 s), coverage was $(100 + 10)/179 = 61\%$. A 0% score indicated that a video was never set into play mode. A score of 100% was the maximum, indicating that every video second had been played at least once. The coverage measure is a proxy for viewing. To record actual viewing, process data such as eye-movement records are needed. This was not possible in the present study.

Commitment was the number of seconds that a video had been set in play mode, expressed as a percentage of the total number of seconds in a video. Unlike coverage, the commitment score could be higher than 100%. A score over 100% revealed that certain video sections had been played at least twice. For example, if a student replayed sections of Clip #1 yielding a total of 214 s of playing for the 179 s video, commitment was $(214/179) \times 100 = 119\%$. Other than coverage, the commitment measure considered replaying or skipping of the videos. Commitment is frequently employed a standard metric for engagement (e.g., Guo et al., 2014; Wistia, 2012). Like coverage, it is a proxy for viewing because one cannot be sure that the user is watching the video when it plays.

Knowledge test. Knowledge was tested with 12 multiple-choice items with four alternatives for each question. The test covered content from all three video clips. An example of a factual item was "What symbol is best used to indicate standard deviation of a sample?" (Alternatives: μ, s, γ, and σ). An example of a conceptual item was "How does the shape of a t-distribution change as the sample size increases?" (Alternatives, it becomes: broader, skewed, more normal, flatter). Because the test included a diverse set of facts and concepts, reliability was expected to be low to moderate. This was indeed what was found (i.e., Cronbach's $\alpha = 0.54$). The original test scores had been converted to the conventional 10-point scale used in the University. To achieve a pass on the test, users needed a score of at least 5.5.

Performance test. The performance test consisted of the obligatory course assignment in the practical in which the students had to conduct a t-test using SPSS. The assignment consisted of two exercises that asked students to work with a dataset in SPSS to solve given problems. Different datasets were randomly distributed to students to prevent them from copying each other's work. The problems in the performance test extended beyond the scope of the videos as they not only related to independent samples t-tests, but also paired samples t-tests and one sample t-tests. For example, one question asked students to test if there was a significant difference in the number of working hours between male and female students. Students were asked to answer questions on hypotheses, alpha values, test statistics, p values and conclusions. In addition, they had to provide the relevant SPSS output. The total assignment was scored out of 10, and students would lose 0.25 points for incorrect or missing responses. The exercises, distribution of datasets, and methodology of grading the assignments were identical as in the previous semester which enabled us to compare cohorts. To achieve a pass on the test, users needed a score of at least 5.5.

5.4. Procedure

All enrolled students received an email stating that a video tutorial had been designed that could serve as a supplementary instructional aid to the course. The message invited students to participate and provided a link to the website with the videos and knowledge test. Students were informed that viewing the videos and completing the test would take about 15 min. In addition, they were told that it would prepare them for the upcoming SPSS practical. They were also informed that their results would not earn them course credits.

All participants received the same communication. The link directed them to either the experimental condition or the control condition. The videos in both conditions could be played multiple times, and were accessible during the knowledge test. The students were told to work independently and to complete their participation within one week. Thereafter, the website was made unavailable and students engaged in the prescribed course practical.

5.5. Data analysis

Before analyzing the data to address the research questions, we conducted a check on the random distribution of the participants' prior knowledge. A prior knowledge score for each participant was obtained from the methodological course that immediately preceded Inferential Statistics. This obligatory course introduced statistics (Theory) and software (SPSS) and provided a test score for each facet. [From 22 participants in Inferential Statistics the Theory and SPSS data were missing.] A comparison between the experimental and control group showed that there was no statistical significant difference for Theory, F (85) < 1, or for SPSS, F (85) = 1.44, n.s. In other words, for prior knowledge the random distribution revealed no difference between conditions.

Regression analyses (Stepwise method) were conducted to explore whether prior knowledge and engagement predicted the scores on the knowledge and performance test. These analyses search for the variables that best predict the outcome. This is done by selecting the predictor that has the highest simple correlation with the outcome. If that variable significantly improves the prediction, it is retained in the model and another predictor (akin a partial correlation) is searched. The exploration stops when no more significant predictors are found (Field, 2013). Conducting several regression analyses on the data involves a degree of multiple testing and hence of chance capitalization. For the reader to assess this risk, we also reported the ANOVA data for the proposed model.

Comparisons within conditions were tested with repeated measures t-tests. Comparisons between conditions were tested with Analysis of Variance (ANOVAs). The ANOVA for the cohort comparison on the performance test reported unequal variances and therefore a Welch test was conducted. Because that test yielded the same significance finding as the ANOVA, we simply presented the F-value.

All comparisons involved two-sided tests with alpha levels of 0.05 for significance. Cohen's (1988) d-statistic was used to indicate the effect size, classified as small for $d = 0.20$, medium for $d = 0.50$, and large for $d = 0.80$.

6. Results

6.1. Engagement data

Table 1 presents the data for *coverage*. This engagement measure represents the proportion of a video that has been set in play-mode. Coverage shows the percentage of the unique seconds in a video that the user may have viewed. Table 1 shows that coverage was very high with a mean score of 98% and mean scores for separate videos that were above 90%. There was no difference between conditions on the mean overall score for coverage, $F(1, 108) = 1.07$, n.s.

A comparison of the mean coverage scores for *demonstrations* and *reviews* (within the experimental condition only) showed a statistically significant effect, $t(53) = 3.48$, $p = 0.001$, $d = 0.53$. The mean coverage score for demonstrations was 98.4% (s.d. 6.39). For reviews this score was 89.9% (s.d. 21.9).

Table 2 presents the data for *commitment*. The commitment measure is a sum score of the play moments for a video divided by its length. Table 2 shows that all mean scores are higher than 100%. This means that generally all videos have been played longer than what would have been needed for a single run. There was no difference between conditions on the mean overall score for commitment, $F(1, 108) = 1.07$, n.s.

A comparison of the mean commitment scores for *demonstrations* and *reviews* (within the experimental condition only) showed a statistically significant effect, $t(53) = 4.15$, $p < 0.001$, $d = 0.54$. The mean commitment score for demonstrations was 111.4% (s.d. 29.4). For reviews this score was 96.3% (s.d. 27.7).

6.2. Knowledge test scores

Table 3 shows the findings for the knowledge test. Comparisons showed no effect of condition, $F < 1$. Twelve students (22%), evenly distributed across conditions, did not achieve a pass score on the test.

An exploratory regression analysis with the possible predictors Theory, SPSS, coverage and commitment revealed two significant factors. The first predictor was *Theory* which explained 19.4% of the variance. *Coverage* further increased the explained variance to a total of 26%. The ANOVA data for the proposed model indicated the presence of a statistically significant fit with the data, $F(2,85) = 14.56$, $p < 0.001$. Both theory ($r = 0.44$) and coverage ($r = 0.18$) were positively related to the knowledge test score.

For the review condition, an exploratory regression analysis with the possible predictors Theory, SPSS, demonstration coverage, demonstration commitment, review coverage, and review commitment revealed two significant factors. The first predictor was again *theory* which explained 13.6% of the variance. *Review commitment* further increased the explained variance to a total of 23.1%. The ANOVA data for the proposed model indicated the presence of a statistically significant fit with the data, $F(2,46) = 6.62$, $p = 0.003$. Both theory ($r = 0.37$) and review commitment ($r = 0.26$) were positively related to the knowledge test score.

6.3. Performance test scores

Table 3 also shows the findings for the performance test. Comparisons showed no effect of condition, $F(1,98) = 2.07$, $p = 0.15$. All students who had submitted their performance test received a pass score. The mean scores from the students who took the performance test in the previous cohort ($M = 8.38$, s.d. 1.46) were significantly and moderately lower than for the participants in this study, $F(1,272) = 13.25$, $p < 0.001$, $d = 0.49$.

An exploratory regression analysis with the possible predictors Theory, SPSS, coverage and commitment revealed no significant factors. Likewise, for the review condition, an exploratory regression analysis with the possible predictors Theory, SPSS, demonstration coverage, demonstration commitment, review coverage, and review commitment revealed no significant factors.

7. Discussion

The videos in the study were designed to fill a gap between theory lessons and the practicals on SPSS. The gap existed as the course delivered theory content and SPSS content separately, with limited overlap between the two. The videos acted as a bridge that

Table 1
Mean coverage[a] (standard deviation) per condition and clip.

	Clip #1 Mean (SD)	Clip #2 Mean (SD)	Clip #3 Mean (SD)	*Total* Mean (SD)
Control ($n = 55$)	98.7 (7.61)	98.9 (6.24)	98.0 (9.26)	*98.5 (7.61)*
Review ($n = 54$)	99.7 (1.95)	97.4 (7.55)	93.9 (17.5)	*97.0 (7.77)*
Total ($n = 109$)	*99.2 (5.58)*	*98.2 (6.93)*	*96.0 (14.6)*	*97.8 (7.69)*

[a] A coverage score of 0% means the whole video is skipped; a coverage score of 100% mean that each unique video second has been played.

Table 2
Mean commitment[a] (standard deviation) per condition and clip.

	Clip #1 Mean (SD)	Clip #2 Mean (SD)	Clip #3 Mean (SD)	*Total* Mean (SD)
Control ($n = 55$)	121.4 (45.2)	110.9 (37.0)	126.1 (61.0)	*119.5 (43.1)*
Review ($n = 54$)	117.6 (37.9)	107.3 (28.7)	110.7 (48.8)	*111.9 (32.6)*
Total ($n = 109$)	*119.5 (41.6)*	*109.1 (33.0)*	*118.5 (55.6)*	*115.7 (38.3)*

[a] A commitment score of 100% indicates the number of play moments equal the length of the video; higher scores indicate repeated plays of sections.

Table 3
Means[a] (standard deviation) on the knowledge test and performance test per condition.

	Knowledge test Mean (SD)	Performance test Mean (SD)
Control ($n = 55$)	6.64 (1.84)	8.84 (0.96)
Review ($n = 54$)	6.94 (1.80)	9.10 (0.82)
Total ($n = 109$)	*6.79 (1.82)*	*8.97 (0.90)*

[a] The maximum test score is 10; a result of 5.5 or higher indicates a pass score.

students could use to connect theory to statistical testing with SPSS.

Early in the study, the decision was made to investigate only videos that had been designed in accordance with the DBT-Model. Of course, this meant the absence of a control condition with non-DBT-based videos. The main reason was that we wanted to construct and investigate videos that were fundamentally well-designed. This meant that the inclusion and expression of design features should heed the large body of research on multimedia, and it further meant a challenge in orchestrating these features. The consideration to give the design of the videos our best effort also fitted perfectly with the real context in which they were to serve.

Participation in the study was voluntary. However, the majority of students chose to partake and the videos were very successful in engaging the students. This can be attested through excellent coverage and commitment results, which indicate the videos were successful at gaining and maintaining the motivation and attention of students. The students played the videos (nearly) completely and occasionally replayed sections of them.

The design of the study precluded a comparison of coverage and commitment data for videos not using a DBT-Model. In other words, the empirical results cannot prove that the model was effective. However, we believe that the videos were optimized by considering processes in observational learning and implementing the associated design guidelines. This provides us with a reasonable assumption that explains the high scores for engagement by students.

To add, the commitment data is lower in the experimental condition and replaying sections is more common during the demonstrations than the reviews. This can be interpreted in two ways. First, we can assume that students did not feel the need to replay the reviews as they were effective at presenting a summary of key points. Alternatively, this could mean that once the students saw that the review was summarizing material that had been presented before, they did not wish to see it presented again. If the experiment had asked students on their opinion regarding the reviews, cross-validation might have been found.

Overall, students achieved moderate scores on the knowledge test. To add, although students had the freedom to consult the videos repeatedly after seeing the questions on the test, re-loading the videos was practically unhandy and may have been done infrequently. In future research, additional procedural controls are needed to investigate the knowledge effect better.

There was no main effect of review condition for knowledge test scores. However, overall regression analysis showed that coverage and theory from the prerequisite course were predictors. The data suggests that greater coverage, and therefore more video information that has potentially been viewed, is associated with higher knowledge scores. In addition, it suggests that students' theoretical knowledge from the related, prerequisite course influenced their viewing behavior and associated knowledge test scores.

Additionally, an interesting outcome was obtained for the regression analysis when the review was distinguished from the demo. Like the overall regression findings, theory was the main predictor, only now the review commitment was also significant factor. Again, this suggests motivation may have played a role on student behavior, and we assume from the data that students who played the review longer are associated with higher knowledge test scores. The finding helps demonstrate the contribution of the reviews. In addition, the finding suggests that students may not be aware of, or sufficiently motivated for using reviews towards their own benefit.

Scores were very high on the performance test for students in both video conditions and regression analysis did not show any predictors. There was a time gap between when students watched the videos and completed the performance test, and all students received the same classroom lesson prior to receiving this test. In addition, the videos instructed students only on one of the three t-tests included in the performance test. This may explain why there were no significant results. However, performance test scores were significantly higher than the previous cohort when students did not have access to the videos. This suggests the important role videos can play in improving statistics instructions in which there is a disparity between theory and practice.

8. Conclusion

This study set out to construct a set of videos that could connect key statistics facts and concepts with their application in statistical tests, including the use of software. For this purpose, a DBT-model was advanced that presented instructional features for supporting the key processes of observational learning proposed by Bandura (1986). The model was adapted from Grossman et al. (2013) to fit the aims and context of software training. The instructional features in the model were derived from, or connected to, existing frameworks on multimedia learning (e.g., Koumi, 2013; Mayer, 2014a; van der Meij & van der Meij, 2013).

The description of each key process in the model was followed by a discussion of features that included evidence from empirical studies of their effect on learning and/or motivation. The method illustrated how these features were implemented to construct the three video clips on conducting a *t*-test in SPSS. The data indicated that the videos yielded high engagement scores, and satisfactory to good scores on a knowledge and performance test. This was seen as tentative support for the proposed DBT-model.

One aspect that has been neglected in the present study concerns the students' opinion about video usage. The technology adoption model (e.g., Davis, 1989; Davis, Bagozzi, & Warshaw, 1989; Joo, Lee, & Ham, 2014) suggests that user's views concerning perceived ease of use and perceived usefulness can play an important role in their willingness to engage with video. Future studies involving video in educational settings might therefore want to complement engagement and effectiveness data with such appraisals.

The study found no main effects for reviews. Only one outcome from the exploratory regression analysis revealed a significant contribution. This contrasts with the positive contributions reported for reviews in two recent studies (van der Meij & van der Meij, 2016a,b). An important difference with these studies is the complexity of task domain. The video in the present study targeted statistics, and was designed to support both conceptual and procedural knowledge development. Another difference is that the present study was conducted in a real context with much less control than in the experimental setting of the other two studies. To add, the statistical power of the present study was limited to one course, with experiment data from one set of students. Confidence in what does and does not work for statistics instruction could be heightened with stronger statistical power. In addition, as one reviewer kindly reminded us, exploratory research as presented here would benefit from pre-registration methods advocated by groups such as the Center for Open Science (COS). With that being said, we hope the current study will inspire future research in the field, and serve as a vehicle to promote better practices. For now, it seems too early to decide whether or not reviews make a substantial contribution to the user's motivation and learning from video.

Ethical standards

Compliance with *Ethical Standards*: A research proposal describing the study has been submitted to the ethics committee of the University who has given approval.

Funding

No *funding* was obtained for this study.

Conflict of interest

The authors declare that they have *no conflict of interest*.

References

Anderson, J. R. (2010). *Cognitive psychology and its implications* (7th ed.). New York, NY: Worth.

Baglin, J., & Da Costa, C. (2014). How do students learn statistical packages? A qualitative study. In H. MacGillivray, B. Phillips, & M. Martin (Eds.), *Topics from Australian conferences on teaching statistics. Springer proceedings in mathematics & statistics* (Vol. 81, pp. 169–187). New York, NY: Springer.

Bandura, A. (1986). *Social foundations of thought and actions: A social cognitive theory*. Englewood Cliffs, NJ: Prentice Hall.

Bandura, A. (1997). *Self-efficacy. The exercise of control*. New York, NY: Freeman and Company.

Berney, S., & Bétrancourt, M. (2016). Does animation enhance learning? A meta-analysis. *Computers & Education, 10*, 150–167.

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

Boucheix, J. M., & Guignard, H. (2005). What animated illustrations conditions can improve technical document comprehension in young students? Format, signaling and control of the presentation. *European Journal of Psychology of Education, 20*(4), 369–388.

Boucheix, J. M., & Lowe, R. (2010). An eye tracking comparison of external pointing cues and internal continuous cues in learning with complex animations. *Learning and Instruction, 20*, 123–135.

Boucheix, J. M., Lowe, R. K., & Bugaiska, A. (2015). Age differences in learning from instructional animations. *Applied Cognitive Psychology, 29*, 524–535.

Bracken, C. C., Pettey, G., Guha, T., & Rubenking, B. E. (2010). Sounding out small screens and telepresence. The impact of audio, screen size and pace. *Journal of Media Psychology, 22*(3), 125–137.

Brucker, B., Scheiter, K., & Gerjets, P. (2014). Learning with dynamic and static visualizations: Realistic details only benefit learners with high visuospatial abilities. *Computers in Human Behavior, 36*, 330–339.

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

Carroll, J. M. (Ed.). (1998). *Minimalism beyond the nurnberg funnel*. Cambridge, MA: MIT Press.

Carroll, J. M., & van der Meij, H. (1998). Ten misconceptions about minimalism. In J. M. Carroll (Ed.), *Minimalism beyond the nurnberg funnel* (pp. 55–90). Cambridge, MA: MIT Press.

Catrambone, R. (1995). Aiding subgoal learning: Effects on transfer. *Journal of Educational Psychology, 87*(1), 5–17.

Catrambone, R. (1998). The subgoal learning model: Creating better examples so that students can solve novel problems. *Journal of Experimental Psychology: General, 127*, 355–376.

Clark, R. E. (2015). Motivating performance: Part 1-diagnosing and solving motivation problems. *Performance Improvement, 54*(8), 33–43.

Clarke, T., Ayres, P., & Sweller, J. (2005). The impact of sequencing and prior knowledge on learning mathematics through spreadsheet applications. *Educational Technology Research & Development, 53*(3), 15–24.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Cross, M. (2010). *The purpose of a film trailer*. Retrieved from http://www.slideshare.net/melaniacross/the-purpose-of-a-film-trailer.

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly, 13*(3), 319–340.

Davis, F. D., Bagozzi, R. P., & Warshaw, P. R. (1989). User acceptance of computer technology: A comparison of two theoretical models. *Management Science, 35*(8), 982–1003.

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

Field, A. (2013). *Discovering statistics using IBM SPSS statistics* (4 ed.). London, UK: Sage.

Finsterwalder, J., Kuppelwieser, V. G., & de Villiers, M. (2012). The effects of film trailers on shaping consumer expectations in the entertainment industry: A qualitative analysis. *Journal of Retailing and Consumer Services, 19*(6), 585–595.

Fulford, C. P. (1992). Systematically designed text enhanced with compressed speech audio. In *Paper presented at the annual meeting of the association for educational communications and technology* (Washington, D.C).

Ginns, P., Martin, A. J., & Marsh, H. W. (2013). Designing instructional text in a conversational style: A meta-analysis. *Educational Psychology Review, 25*, 445–472.

van Gog, T., Kester, L., & Paas, F. (2011). Effects of worked examples, example-problem, and problem-example pairs on novices' learning. *Contemporary Educational Psychology, 36*, 212–218.

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

Guo, P. J., Kim, J., & Rubin, R. (2014, March 4–5). How video production affects student engagement: An empirical study of MOOC videos. In *Paper presented at the L@S '14* (Atlanta, GA).

Hasler, B. S., Kersten, B., & Sweller, J. (2007). Learner control, cognitive load and instructional animation. *Applied Cognitive Psychology, 21*, 713–729.

Hassanabadi, H., Robatjazi, E. S., & Savoji, A. P. (2011). Cognitive consequences of segmentation and modality methods in learning from instructional animations. *Procedia Social and Behavioral Sciences, 30*, 1481–1487.

Helsdingen, A., van Gog, T., & van Merriënboer, J. J. G. (2011). The effects of practice schedule and critical thinking prompts on learning and transfer of a complex judgment task. *Journal of Educational Psychology, 103*(2), 383–398.

Höffler, T. N., & Schwartz, R. N. (2011). Effects of pacing and cognitive style across dynamic and non-dynamic representations. *Computers & Education, 57*, 1716–1726.

ten Hove, P., & van der Meij, H. (2015). Like it or not. What characterizes YouTube's more popular videos? *Technical Communication, 62*(1), 48–62.

Joo, Y. J., Lee, H. W., & Ham, Y. (2014). Integrating user interface and personal innovativeness into the TAM for mobile learning in cyber university. *Journal of Computing in Higher Education, 26*(2), 143–158. http://dx.doi.org/10.1007/s12528-014-9081-2.

Kalyuga, S. (2007). Expertise reversal effect and its implications for learner-tailored instruction. *Educational Psychology Review, 19*, 509–539.

Kämpfe, J., Sedlmeier, P., & Renkewitz, F. (2010). The impact of background music on adult listeners: A meta-analysis. *Psychology of Music, 39*(4), 424–448.

Kay, R. H. (2014). Developing a framework for creating effective instructional video podcasts. *International Journal of Emerging Technologies in Learning, 9*(1), 22–30.

Kay, R. H., & Kletskin, I. (2012). Evaluating the use of problem-based video podcasts to teach mathematics in higher education. *Computers & Education, 59*, 619–627.

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.

Kosslyn, S. M., Kievit, R. A., Russell, A. G., & Shephard, J. M. (2012). PowerPoint presentation flaws and failures: A psychological analysis. *Frontiers in Psychology, 3*, 1–22.

Koumi, J. (2013). Pedagogic design guidelines for multimedia materials: A call for collaboration between practitioners and researchers. *Journal of Visual Literacy, 32*(2), 85–114.

Kriz, S., & Hegarty, M. (2007). Top-down and bottom-up influences on learning from animations. *International Journal of Human Computer Studies, 65*, 911–930.

Lang, A., Park, B., Sanders-Jackson, A. N., Wilson, B. D., & Wang, Z. (2007). Cognition and emotion in tv message processing: How valence, arousing content, structural complexity, and information density affect the availability of cognitive resources. *Media Psychology, 10*, 317–2007.

Lemarié, J., Lorch, R. F., Eyrolle, H., & Virbel, J. (2008). Sara: A text-based and reader-based theory of signaling. *Educational Psychologist, 43*(1), 27–48.

Leppink, J., Paas, F., van Gog, T., van der Vleuten, C. P. M., & van Merriënboer, J. J. G. (2014). Effects of pairs of problems and examples on task performance and different types of cognitive load. *Learning and Instruction, 30*, 32–42.

Lloyd, S. A., & Robertson, C. L. (2012). Screencast tutorials enhance student learning of statistics. *Teaching of Psychology, 39*(1), 67–71.

Lowe, R., Schnotz, W., & Rasch, T. (2011). Aligning affordances of graphics with learning task requirements. *Applied Cognitive Psychology, 25*, 452–459.

Lusk, D. L., Evans, A. D., Jeffrey, T. R., Palmer, K. R., Wikstrom, C. S., & Doolittle, P. E. (2009). Multimedia learning and individual differences: Mediating the effects of working memory capacity with segmentation. *British Journal of Educational Technology, 40*(4), 636–651.

Margulieux, L., Guzdial, M., & Catrambone, R. (2012). Subgoal-labeled instructional material improves performance and transfer in learning to develop mobile applications. In *Paper presented at the 9th international conference on international computing education research (ICER'12)* (Auckland, New Zealand).

Mayer, R. E. (2011). Instruction based on visualization. In R. E. Mayer, & P. A. Alexander (Eds.), *Handbook of research on learning and instruction* (pp. 427–445). New York, NY: Routledge.

Mayer, R. E. (2014a). *The Cambridge handbook of multimedia learning* (2nd ed.). New York, NY: Cambridge University Press.
Mayer, R. E. (2014b). Cognitive theory of multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2 ed, pp. 43–71). New York, NY: Cambridge University Press.
Mayer, R. E., & Pilegard, C. (2014). Principles for managing essential processing in multimedia learning: Segmenting, pre-training, and modality principles. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed., pp. 316–344). New York, NY: Cambridge University Press.
van der Meij, H. (2014). Developing and testing a video tutorial for software learning. *Technical Communication, 61*(2), 110–122.
van der Meij, H., Blijleven, P., & Jansen, L. (2003). What makes up a procedure? In M. J. Albers, & B. Mazur (Eds.), *Content & complexity. Information design in technical communication* (pp. 129–186). Mahwah, NJ: Erlbaum.
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* (pp. 19–53). Cambridge, MA: MIT Press.
van der Meij, H., & Gellevij, M. R. M. (2004). The four components of a procedure. *IEEE Transactions on Professional Communication, 47*(1), 5–14.
van der Meij, H., & van der Meij, J. (2013). Eight guidelines for the design of instructional videos for software training. *Technical Communication, 60*(3), 205–228.
van der Meij, H., & van der Meij, J. (2014). A comparison of paper-based and video tutorials for software learning. *Computers & Education, 78*, 150–159.
van der Meij, J., & van der Meij, H. (2015). A test on the design of a video tutorial for software training. *Journal of Computer-Assisted-Learning, 31*(2), 116–132.
van der Meij, H., & van der Meij, J. (2016a). The effects of reviews in video tutorials. *Journal of Computer Assisted Learning, 32*, 332–344.
van der Meij, H., & van der Meij, J. (2016b). Demonstration-Based Training (DBT) for the design of a video tutorial for software instructions. *Instructional Science, 44*, 527–542.
Merkt, M., & Schwan, S. (2014). How does interactivity in videos affect task performance? *Computers in Human Behavior, 31*, 172–181.
Merkt, M., Weigand, S., Heier, A., & Schwan, S. (2011). Learning with videos vs. learning with print: The role of interactive features. *Learning and Instruction, 21*, 687–704.
Meyer, K., Rasch, T., & Schnotz, W. (2010). Effects of animation's speed of presentation on perceptual processing and learning. *Learning and Instruction, 20*, 136–145.
Pintrich, P. R., & Schunk, D. H. (2002). *Motivation in education. Theory, research, and applications* (2nd ed.). Upper Saddle River, NJ: Merrill Prentice Hall.
Pollock, E., Chandler, P., & Sweller, J. (2002). Assimilating complex information. *Learning and Instruction, 12*, 61–86.
Reichelt, M., Kämmerer, F., Niegemann, H. M., & Zander, S. (2014). Talk to me personally: Personalization of language style in computer-based learning. *Computers in Human Behavior, 35*, 199–210.
Renkl, A. (2014). Toward an instructionally oriented theory of example-based learning. *Cognitive Science, 38*, 1–37.
Richter, J., Scheiter, K., & Eitel, A. (2015). Signaling text-picture relations in multimedia learning: A comprehensive meta-analysis. *Educational Research Review, 17*, 19–36.
Rosen, M. A., Salas, E., Pavlas, D., Jensen, R., Fu, D., & Lampton, D. (2010). Demonstration-based training: A review of instructional features. *Human Factors, 52*(5), 596–609.
Schittek Janda, M., Botticelli, T. A., Mattheos, N., Nebel, D., Wagner, A., Nattestad, A., & Attström, R. (2005). Computer-mediated instructional video: A randomised controlled trial comparing a sequential and a segmented instructional video in surgical hand wash. *European Journal of Dental Education, 9*, 53–58.
Schmidt, R. A., & Bjork, R. A. (1992). New conceptualizations of practice: Common principles in three paradigms suggest new concepts for training. *Psychological Science, 3*, 207–217.
Schreiber, B. E., Fukuta, J., & Gordon, F. (2010). Live lecture versus video podcast in undergraduate medical education: A randomised controlled trial. *BMC Medical Education, 10*(1), 1–6.
Schwan, S., Garsoffky, B., & Hesse, F. W. (2000). Do film cuts facilitate the perceptual and cognitive organization of activity sequences? *Memory & Cognition, 28*, 214–223.
Schwan, S., & Riempp, R. (2004). The cognitive benefit of interactive videos: Learning to tie nautical knots. *Learning and Instruction, 14*, 293–305.
Shneiderman, B., Plaisant, C., Cohen, M., & Jacobs, S. (2010). *Designing the user interface: Strategies for effective human-computer interaction* (6 ed.). Boston (NY): Pearson.
Simonds, B. K., Meyer, K. R., Quinlan, M. M., & Hunt, S. K. (2006). Effects of instructor speech rate on student affective learning, recall, and perceptions of nonverbal immediacy, credibility, and clarity. *Communication Research Reports, 23*(3), 187–197.
Smith, E. E., & Kosslyn, S. M. (2007). *Cognitive psychology: Mind and brain* (1st ed.). Harlow, UK: Pearson.
Spanjers, I. A. E., van Gog, T., Wouters, P., & van Merriënboer, J. J. G. (2012). Explaining the segmentation effect in learning from animations: The role of pausing and temporal cueing. *Computers & Education, 59*, 274–280.
Spanjers, I. A. E., Wouters, P., van Gog, T., & van Merriënboer, J. J. G. (2011). An expertise reversal effect of segmentation in learning from animated worked-out examples. *Computers in Human Behavior, 27*, 46–52.
Stiller, K. D., Freitag, A., Zinnbauer, P., & Freitag, C. (2009). How pacing of multimedia instructions can influence modality effects: A case of superiority of visual texts. *Australasian Journal of Educational Technology & Society, 25*(2), 184–203.
Strobin, A. A., Hunt, J. B., Spencer, F. J., & Hunt, T. G. (2015). The role of music in motion picture advertising and theatrical trailers: Altering music to modify emotional response and genre expectations. *Academy of Marketing Studies Journal, 19*(3), R244–R260.
Wagoner, B. (2008). Narrative form and content in remembering. *Integrative Psychological Behavior, 42*, 315–323.
Wistia. (2012, 7 May). *Does length matter?*. Retrieved from http://wistia.com/blog/does-length-matter-it-does-for-video-2k12-edition.
Witteman, M. J., & Segers, E. (2010). The modality effect tested in children in a user-paced multimedia environment. *Journal of Computer Assisted Learning, 26*, 132–142.
Yue, C. L., Bjork, E. L., & Bjork, R. A. (2013). Reducing verbal redundancy in multimedia learning: An undesired desirable difficulty? *Journal of Educational Psychology, 105*(2), 266–277.
Zacks, J. M., Speer, N. K., Swallow, K. M., Braver, T. S., & Reynolds, J. R. (2007). Event perception: A mind-brain perspective. *Psychological Bulletin, 133*(2), 273–293.