ROLE OF INTERACTIVE TASKS AND A COMPREHENSION-ORIENTED DESIGN IN INSTRUCTIONAL VIDEOS: RESULTS OF A PILOT STUDY

Sina Wetzel, Matthias Ludwig

Goethe University Frankfurt; wetzel@math.uni-frankfurt.de, ludwig@math.uni-frankfurt.de

Instructional mathematics videos that can be found e.g. on YouTube are very popular with students as support for doing homework or learning for an exam. However, from a didactical point of view, these videos can often be criticised for only conveying the superficial use of procedures and for leaving students as passive recipients of knowledge. In a randomised experimental study, we aim to find out how interactive tasks and a comprehension-oriented design can support not only the acquisition of procedural but also of conceptual knowledge. This paper reports on the results of a first pilot study with n = 85 participants that will be repeated in summer 2023 in a larger scale. For the acquisition of procedural knowledge, no differences between groups can be reported. For the acquisition of conceptual knowledge, interactive tasks seem to be favourable.

Keywords: Instructional mathematics videos, interactive tasks, procedural knowledge, conceptual knowledge, systems of linear equations

INTRODUCTION

Digital technology plays an important role in students' everyday lives and also gains more and more relevance in educational practice. In mathematics educations, "[t]echnology no longer refers solely to computers, mathematical software (e.g., dynamic geometry software platforms) or graphing calculators" (Moore-Russo et al., 2015, p. 108) but encompasses many different kinds of digital tools such as mobile apps and other media. One type of such tools are instructional videos, sometimes also referred to as educational videos or explanation videos, which are available as free learning resources on YouTube and other platforms. Teachers can integrate these videos into their educational practice; however, students also use these videos on their own outside of lessons to prepare for an exam or as support when completing homework assignments (Bersch et al., 2020). Fiorella (2021) defines instructional videos as "dynamic audiovisual presentations in which an instructor delivers oral explanations while presenting corresponding visual information" (p. 487). This definition mainly focuses on the simultaneous delivery of content in an auditive and visual way which is one of the main advantages attributed to instructional videos according to the cognitive theory of multimedia learning (Mayer, 2022). However, what this definition does not specify is who delivers the content and, even more importantly, how. Another definition by Wolf (2018, p. 4) even focuses on the fact that such videos are usually produced by amateurs who are neither professional media content producers nor designated experts on the topic of the video. In combination with the fact that there are usually no control mechanisms regarding the correctness and didactical substance, the quality of available videos varies. One of the main critical points usually discussed in the rather young research field concerned with instructional mathematics videos is that many of these videos only convey the use of procedures instead of also focusing on concepts and relationships (see e.g. Korntreff & Prediger, 2022; Lobato et al., 2019). Another problem can be the way in which students watch instructional videos. If students are only passive recipients of knowledge and not actively engage with them, they will not be able to actively construct meaningful mathematics knowledge (Bersch et al., 2020). However, as there are also many positive aspects of instructional videos such as the possibility to learn with them anywhere and anytime at one's own individual pace choosing the topics and instructors fitting the own needs and preferences (see e.g. Lobato et al., 2019; Wetzel & Ludwig, 2021a), it is important to dedicate research on how to overcome these problems or at least reduce their impact. In this paper, we present the results of a first study on the effect of interactive tasks and a comprehension-oriented design on the procedural and conceptual knowledge acquired through watching a video on the topic of systems of linear equations.

THEORETICAL FRAMEWORK AND RESEARCH QUESTIONS

It seems desirable to create videos that not only teach students the superficial use and application of procedures (also referred to as procedural knowledge) but also convey conceptual knowledge. Conceptual knowledge can be defined as the "explicit or implicit understanding of the principles that govern a domain and of the interrelations between pieces of knowledge in a domain" (Rittle-Johnson & Alibali, 1999, p. 175). However, to convey conceptual knowledge, videos must actually contain corresponding "concept elements" which many of the existing videos on YouTube do not do (Korntreff & Prediger, 2022). Kulgemeyer (2020) argues that for acquiring conceptual knowledge that can be transferred to new learning situations, further learning activities are needed such as a follow-up task at the end of or right after the video. However, such a learning activity cannot only be situated at the end of or after the completion of a video but also during it, e.g. by the means of interactive tasks (Wetzel & Ludwig, 2021b). Interactive tasks during a video can also potentially increase students' engagement when watching the video which a task at the end or after the video cannot accomplish. To our knowledge, there are no empirical results regarding the acquisition of conceptual knowledge with instructional mathematics videos with the target group of high school students, neither for videos with nor without interactive tasks. We thus want to examine the acquisition of conceptual knowledge more closely, also when compared to procedural knowledge, and the influence of interactive tasks on both knowledge types. In order that concepts can be learned at all, videos must be designed in a comprehension-oriented way. As our research interest are videos viewed by the students on their own which are usually videos on a topic the students already know for homework completion or exam preparation, we also use videos with the purpose of repetition. We want to answer two main research questions:

RQ1: How does a *comprehension-oriented design* of an instructional mathematics video, viewed with the purpose of repetition, influence the acquisition of procedural and conceptual knowledge?

RQ2: How do *interactive tasks* in an instructional mathematics video, viewed with the purpose of repetition, influence the acquisition of procedural and conceptual knowledge?

Regarding RQ1, it is likely that a comprehension-oriented design will positively influence the acquisition of conceptual knowledge, if conceptual knowledge can be acquired by means of a video alone. It is unclear, how interactive tasks will influence the acquisition of the two knowledge types, however we assume that students will have a higher cognitive engagement and thus have higher test scores, especially regarding the acquisition of conceptual knowledge which needs additional learning activities according to Kulgemeyer (2020).

METHODS

Study setup

To answer the research questions, we conduct a randomised experimental study in a pretest posttest design with four different treatment groups which will each view a different video:

- Group 1: Video is *not* comprehension-oriented[1] and does *not* contain interactive tasks
- Group 2: Video is *not* comprehension-oriented and contains interactive tasks
- Group 3: Video is comprehension-oriented and does *not* contain interactive tasks
- Group 4: Video is comprehension-oriented and contains interactive tasks

The whole study is conducted in one sitting. At the beginning, each student receives a code snippet randomly assigning them to one of the four treatment groups. Students then write a pretest in 25 minutes, watch the videos on tablets with headphones at their own pace in a maximum of 20 minutes and write a posttest in 25 minutes. A first pilot study was conducted in January and February 2023 on the results of which this paper reports. The main study with a larger sample size will be conducted in summer 2023.

Creation of videos

To create the videos, we relied on criteria for good instructional videos in general and for mathematics videos in particular. Many general criteria are linked to the cognitive theory of multimedia learning which includes recommendations such as no distracting elements, a corresponding presentation of visual and auditive content and a visual highlighting of the most important points in the video (Mayer, 2022; Brame, 2016). Then there are criteria related to the length of videos. One recommendation by Guo et al. (2014) based on the analysis of about seven million video watching sessions is that videos should not be longer than six minutes to guarantee a maximum level of engagement. From a didactical point of view, criteria often mentioned are the use of a positive language when talking about mathematics (Bersch et al., 2020), the use of an interesting introductory example (Bersch et al. 2020; Kulgemeyer, 2020), a discussion of necessary pre-knowledge (Ratnayake et al., 2019) and common misconceptions and a summary of the content in the end (Kulgemeyer, 2020).

As a topic for the videos, we decided on systems of linear equations which are usually taught in grade 9 in Germany with two equations and two variables. We chose this topic, since there are different algorithmic rule-based approaches to determine the set of solutions (e.g. method of substitution and method of elimination) so the topic can be taught focusing on procedures alone but also there are many opportunities to teach concepts such as the meaning of a certain result for the set of solutions. The whole topic of systems of linear equation is too large to be covered by just one video. Therefore, the videos focus on the subtopic “determining and interpreting the set of solutions”. To determine the set of solutions, there are different procedures available. Again, all of them would be too much content so we decided to focus on how to solve systems of linear equations graphically and on the substitution method. For the interpretation part, we focus on the relation between the symbolic and graphical representation of systems of linear equations and how inferences can be made from the set of solutions on the corresponding system of linear equations and vice versa.

On the technical side, the videos were created as screencasts of animated PowerPoint presentations. At first, we created videos of the type “Khan-style tablet drawing” (Guo et al., 2014, p. 42), however, it took longer to write equations and calculations in handwriting which was problematic related to the recommendation of a maximum video length of six minutes by the same authors. Guo et al. (2014) propose to divide a video in several videos in such a case which might make sense for a YouTube channel owner but was not applicable to our study were the video is meant to be watched in one sitting. This aspect is problematic in general and requires further research if a video

[1] Not comprehension-oriented means that these videos focus mainly on procedures

is meant to cover more than just a procedure while at the same time following the design guidelines recommended in literature. To keep our videos at a maximum length of ten minutes, we still had to drop some criteria such as the discussion of common misconceptions and necessary pre-knowledge and we had to change the style to animated PowerPoint presentations.

For the comprehension-oriented video, we added an extensive discussion of the relation between the geometrical location of two lines and the corresponding system of linear equations and set of solutions. For the video with focus on procedures, we only quickly mentioned this aspect and instead dedicated time to a further example for the method of substitution. The interactive tasks were added after the videos had been created using the software H5P (https://h5p.org) with the content type "interactive video" which allows to add different interactive elements to an input video. We used the interactive elements "Single Choice set", "Text" and "Fill in the Blanks" (the latter for the input of numbers, e.g. "Find the value of x.") and created six short tasks. As the purpose of this study is not to find out what kind of interactive tasks are most effective, we decided to use tasks with the sole purpose of activating students. It is central to the study design that the tasks do not provide extra learning opportunities and thus more time on task, therefore we only used tasks and questions which were answered later in the video anyway.

Test instrument

To our knowledge, no publicly available test exists targeting systems of linear equation divided in items for procedural and conceptual knowledge. Furthermore, the study setup in which only 25 minutes are available for pretest and posttest each requires a test than can measure these constructs reliably with few items. The study thus also aims at piloting the developed test instrument and to analyse the items within it. Scanning several textbooks for grade 9, we identified typical types of tasks focussing on systems of linear equations. Some task types could not be considered for the test if they required both procedural and conceptual knowledge to be completed. An example are word problems since these tasks usually require conceptual knowledge for understanding the problem and procedural knowledge to solve the resulting equations. Based on these preliminary considerations, we created 34 tasks which were administered to five professors for the didactics of mathematics for an expert review procedure. Each item was rated on a six-point linear scale with the left end representing tasks only requiring procedural knowledge and the right end tasks only requiring conceptual knowledge in reference to Rittle-Johnson and Alibali (1999) stating that "[t]hese two types of knowledge lie on a continuum and cannot always be separated; however, the two ends of the continuum represent two different types of knowledge" (p. 175). Items with a mean result between 1 and 3 were then considered as requiring rather procedural knowledge while items with a mean value between 4 and 6 as requiring rather conceptual knowledge. Based on this, we accepted items with a mean within these two limits if their standard deviation is less than 1. Applying this rule, seven items were eliminated, nine were identified as procedural and 18 were identified as conceptual knowledge items. Twelve of the 18 conceptual items were chosen for the test and the nine procedural items were supplemented by three further items with similar assignments but different numbers. Procedural items include arithmetical problems such as algorithmically determining the set of solutions, checking whether a value pair is a solution or ordering the steps of the substitution method. Conceptual items mostly rely on the relation between the symbolic and the graphical interpretation of systems of linear equations and the set of solutions such as multiple choice questions were all true statements for a certain context must be marked, or a given linear equation where the student must find another linear equation such that the set of solution meets certain conditions or in the form of reasoning tasks. Both pretest and posttest consist of twelve items (six items rather requiring procedural and six rather requiring conceptual knowledge).

However, not all participants are administered the same items. Three procedural and three conceptual items, respectively, form a block of items. Thus, there are four such blocks in total (A1, A2, B1, B2), which can be combined into one test. Following a multi-matrix design to avoid order effects, one test booklet consisted either of item block A1A2, A2B1, B1B2 or B2A1. The posttest each participant receives depends on the pretest so that each participants get administered all 24 items.

Participants

The pilot study was conducted in January and February 2023 in one German high school in three grade 9 classes and one grade 10 class. A total of n = 85 students participated in the study. Each student was given a code randomly at the beginning assigning them to one of the four treatment groups and one of the test booklets. Due to missing ill students in each class, not all codes were distributed resulting in $n_1 = 24$ participants in group 1, $n_2 = 21$ participants in group 2, $n_3 = 18$ participants in group 3 and $n_4 = 22$ participants in group 4.

Data analysis

Items were rated dichotomously by one rater and raw scores were processed further in R using the TAM package. Due to the use of different test booklets, item scores were equated first using the method of joint calibration (Wu et al., 2016). With the equated data set, a two-dimensional IRT model analysis (first dimension: procedural items, second dimension: conceptual items) was conducted using the results of the pretest and posttest as if they were results of different participants to estimate the item difficulty parameters. In a second step, a four-dimensional IRT model analysis (dimensions: procedural items pretest, procedural items posttest, conceptual items pretest, conceptual items postest) was conducted to estimate the person ability parameters using the item difficulty parameters from the first analysis as input. Based on these person ability parameters, classical test theory tests were conducted. To answer RQ1, group 1 and group 2 are analysed together as group NCO (not comprehension-oriented) and groups 3 and 4 are analysed together as group CO (comprehension-oriented). Similarly, for RQ2, group 1 and 3 together form group NIT (no interactive tasks) and group 2 and 4 form group IT (interactive tasks).

RESULTS

A fit analysis of the 24 test items yielded infit-t values between -1.45 and 1.61 which is inside the range of [-1.96, 1.96] so there is no indication that the items do not fit the model (Wu et al., 2016). Therefore, all items were used for the estimation of person ability parameters.

In the pretest, the mean person ability parameter is -0.38 (SD = 0.86) for the procedural items and -0.52 (SD = 0.95) for the conceptual items. A one-way ANOVA analysis was performed to determine whether pretest scores are comparable for the four treatment groups. There was no statistically significant difference in the pretest scores between groups (($F(1, 82) = 0.66$, $p = 0.418$) for conceptual items. For procedural items, however, there was a statistically significant small difference between at least two groups (($F(1, 82) = 4.56$, $p = 0.036$). We therefore performed two further ANOVAs for procedural items to compare groups NCO and CO ($F(1, 82) = 7.19$, $p = 0.009$) as well as groups NIT and IT ($F(1, 82) = 0.10$, $p = 0.748$). As there is a statistical difference in procedural items scores between groups NCO and CO, we only report on results for groups NIT and IT for procedural items.

A summary of the changes of person ability parameters from pretest to posttest can be found in Table 1. In all groups, the mean person ability parameter increased. A one sample t-test was computed for all reported increases in Table 1 to determine whether they are different from zero. All reported increases differ statistically significant from zero ($p < 0.01$).

Groups	NCO	CO	NIT	IT	All
Procedural items	-	-	**0.84** (0.48)	**0.67** (0.48)	**0.76** (0.46)
Conceptual items	**1.37** (0.44)	**1.42** (0.50)	**1.24** (0.47)	**1.55** (0.42)	**1.39** (0.47)

Table 1. Mean and standard deviation (in brackets) for differences of person ability parameter from pretest to posttest

To answer our first research question, we compare person ability parameter increases for conceptual items for groups NCO and CO. Even though the group CO has a slightly higher mean increase than the group NCO (1.42 compared to 1.37), an ANOVA analysis yielded that this difference is not statistically different ($F(1, 82) = 0.18$, $p = 0.674$). We can therefore not conclusively give an answer to our research question based on this data.

For the second research question, we can compare both, differences for procedural and conceptual items, between the groups NIT and IT. Contrarily to what we expected, the increase of person ability parameters from pretest to posttest is higher for videos with no interactive tasks ($M = 0.84$) than for videos with interactive tasks ($M = 0.67$) for procedural items. However, this difference is not statistically significant based on the results of an ANOVA ($F(1, 82) = 2.81$, $p = 0.098$). For conceptual items, we can report a higher increase for group IT ($M = 1.55$) than group NIT ($M = 1.24$). Again, an ANOVA was performed and this difference was statistically significant ($F(1, 82) = 10.82$, $p = 0.001$). This effect is also visualised in Figure 1. So, to answer our second research question, we can state that interactive tasks in instructional mathematics videos can have a positive effect on the acquisition of conceptual knowledge.

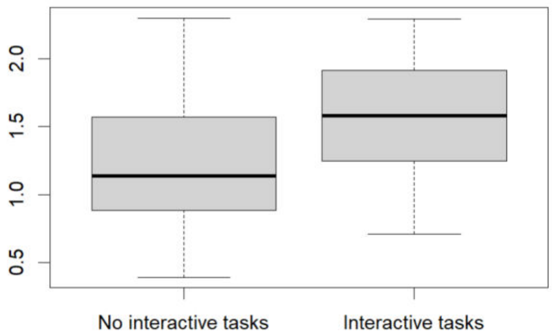

Figure 1. Comparison of differences of person ability parameters for conceptual items from pretest to posttest of the groups NIT and IT

DISCUSSION

For procedural items, we cannot report any statistically significant differences between groups NIT and IT (and we had to exclude NCO and CO from the analysis, see results section). One might expect that interactive tasks which are meant to lead to more cognitive engagement also lead to more procedural knowledge. Especially, since interactive tasks in our study focused mainly on procedures, for example, in an equation with just one variable x, participants were prompted to calculate and enter the correct value for x. However, at least for our small sample size, we cannot report any effects on procedural knowledge. That leads to the question whether for the acquisition of procedural knowledge the interactive design of the video might not be as relevant. The larger scale main study in summer 2023 will hopefully contribute to the answer of this question.

For conceptual knowledge, the first important result to report is that all four groups significantly improved from pretest to posttest. This is also interesting considering that participants in group NCO received a video that focused mainly on procedures. There is not even a statistically significant difference between groups NCO and CO. However, mathematical relations were not omitted completely from the videos in group NCO but rather mentioned shortly without any detail. It would be an interesting insight if such short mentions can already lead to the acquisition of conceptual knowledge. In the main study, the explicit relationship between the video content and the results must therefore be examined more closely. It is likely, however, that the increase in conceptual knowledge based on the videos in group NCO could not be reproduced in a follow-up study. Whether this is the case for group CO is another interesting question for future research.

Regarding the difference between the groups NIT and IT, the groups with interactive tasks had a significantly higher increase from pretest to posttest for conceptual items than the group without interactive tasks. This is in line with Kulgemeyer (2020) stating that follow-up learning activities are necessary for a conceptual knowledge increase and our assumption that interactive tasks during the video can have the same function. However, even if interactive tasks lead to more conceptual knowledge, they were not a prerequisite, as all groups increased their conceptual knowledge.

CONCLUSION AND OUTLOOK

The biggest limiting factor of the presented results is the small sample size. Therefore, we cannot conclusively answer our research questions and need to rely on the results of the main study which will be conducted in summer 2023. However, we can still report that interactive tasks seem to positively influence the acquisition of conceptual knowledge.

In the main study, there will also be an extensive analysis of a corresponding questionnaire with items on topics such as engagement during the video which will also give insights on the participants' level of cognitive activation reached through the interactive tasks. Furthermore, data will be collected regarding the way the participants interact with the video such as pausing, rewinding, or jumping forwards and how much time was spent on each interactive task (for group IT) and if the answers were correct. These additional data will provide more information based on which the research questions can be answered.

One general observation is the complex interplay between the length of a video and its didactical value which proved to be a real practical challenge for us. Further research must be dedicated to investigate the impact didactical criteria and the video length have on student learning and how these two factors interrelate.

REFERENCES

Bersch, S., Merkel, A., Oldenburg, R., & Weckerle, M. (2020). Erklärvideos: Chancen und Risiken - Zwischen fachlicher Korrektheit und didaktischen Zielen. [Explanation videos: Opportunities and risks – Between technical correctness and didactical goals]. *Mitteilungen der Gesellschaft für Didaktik der Mathematik, 46*(109), 58-63.

Brame, C. J. (2016). Effective educational videos: Principles and guidelines for maximizing student learning from video content. *CBE—Life Sciences Education, 15*(4), 1-6.

Fiorella, L. (2021). Multimedia Learning with Instructional Video. In R. Mayer & L. Fiorella (Eds.), *The Cambridge Handbook of Multimedia Learning* (pp. 487-497). Cambridge University Press. https://doi.org/10.1017/9781108894333.050

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

Korntreff, S., & Prediger, S. (2022). Conceptual Learning Opportunities of Instructional YouTube Videos – Conceptualization and Analysis for the Case of Algebraic Concepts. *Journal für Mathematik-Didaktik, 43*, 281-310.

Kulgemeyer, C. (2020). A framework of effective science explanation videos informed by criteria for instructional explanations. *Research in Science Education, 50*(6), 2441-2462.

Lobato, J., Walters, C. D., Walker, C., & Voigt, M. (2019). How Do Learners Approach Dialogic, On-Line Mathematics Videos? *Digital Experiences in Mathematics Education, 5*(1), 1–35. https://doi.org/10.1007/s40751-018-0043-6

Mayer, R. E. (2022). Cognitive Theory of Multimedia Learning. In R. E. Mayer & L. Fiorella (Eds.), *The Cambridge Handbook of Multimedia Learning* (pp. 57-72). Cambridge University Press. https://doi.org/10.1017/9781108894333.050

Moore-Russo, D., Diletti, J., Strzelec, J., Reeb, C., Schillace, J., Martin, A., Arabeyyat, T, Prabucki, K. & Scanlon, S. (2015). A study of how angry birds has been used in mathematics education. *Digital Experiences in Mathematics Education, 1*, 107-132.

Ratnayake, I., Bruder, R., Johlke, F., & Feldt-Caesar, N. (2019). Quality criteria for teachers to choose video tutorials for different learning situations. In *EDULEARN19 Proceedings* (pp. 3669-3674). IATED.

Rittle-Johnson, B., & Alibali, M. W. (1999). Conceptual and procedural knowledge of mathematics: Does one lead to the other? *Journal of educational psychology, 91*(1), 175-189.

Wetzel, S., & Ludwig, M. (2021a). Categorising perceived advantages of instructional mathematics videos. In M. Inprasitha, N. Changsri & N. Boonsena (Eds.), *Proceedings of the 44th Conference of the International Group for the Psychology of Mathematics Education*, Vol. 4 (pp. 273-281).

Wetzel, S. & Ludwig, M. (2021b). Mathematische Erklärvideos interaktiv gestalten: Ansätze für Lehre und Forschung. [Designing instructional mathematics videos in an interactive way: Ideas for teaching and research]. In: Kerstin Hein, Cathleen Heil, Silke Ruwisch & Susanne Prediger (Eds.). *Beiträge zum Mathematikunterricht 2021* (pp. 181-184). WTM Verlag.

Wolf, K. D. (2018). Video statt Lehrkraft? Erklärvideos als didaktisches Element im Unterricht. [Video instead of the teacher? Explanation videos as didactical element in lessons]. *Computer+Unterricht, 109*, 4–7.

Wu, M., Tam, H. P. & Jen, T.-H. (2016). *Educational Measurement for Applied Researchers: Theory into Practice*. Springer.