



A Rating Framework for the Quality of Video Explanations

Malte Ring¹ · Taiga Brahm¹

Accepted: 23 November 2022
© The Author(s) 2022

Abstract

More and more teachers create video explanations for their instruction. Whether or not they are effective for learning depends on the videos' instructional quality. Reliable measures to assess the quality of video explanations, however, are still rare, especially for videos created by (preservice) teachers. We developed such a measure in a two-step process: First, the categories were theoretically derived. Second, a coding manual was developed and used with 36 videos, which were created by preservice teachers during a university seminar. The resulting framework, which can be used as a coding manual for future research, consists of twelve criteria in five different categories: video content, learner orientation, representation and design, language, and process structure. With this framework, we contribute a reliable measure to evaluate the quality of existing videos. In practice, teachers can also use this measure as a guideline when creating or choosing video explanations for the classroom.

Keywords Video explanation · Explanatory videos · Instructional videos · Instructional quality · Teacher-created videos · Economic education · Instrument development

1 Introduction

With the rise of video platforms and the simplified availability of digital tools, an increasing number of teachers use videos for their teaching. While many teachers take existing videos from the internet, more and more teachers also create short video explanations as they can be specifically tailored to their teaching and, consequently, to their students' learning goals (Jaekel et al., 2021).

Scholars have used a variety of labels for videos used in the instructional context, such as instructional videos, demonstration videos, Kahn academy videos, classroom videos or video explanations (Köse et al., 2021). To highlight their explanatory character, we use the term video explanations to describe videos that combine spoken language and visualizations to explain a topic, a principle, or a process to learners. Thus, these videos merge features of traditional instructional explanations (in the classroom) with the advantages of a multimedia learning environment (Kulgemeyer, 2018a).

✉ Malte Ring
malte.ring@uni-tuebingen.de

¹ Economic Education, University of Tuebingen, Melanchthonstraße 30, 72074 Tübingen, Germany

As the quality of video explanations is relevant for students' learning outcomes (Kulgemeyer, 2018a), it is important to establish a comprehensive framework for evaluating the quality which can be used by researchers and practitioners to make an informed decision. Different researchers already summarized recommendations for effective video instruction which mostly rely on multimedia principles and often focus on videos in the STEM domains (e.g., Brame, 2016; Kay, 2014). Another strand of research analyzes video quality based on research regarding explanations in the classroom (Kulgemeyer, 2018a).

Using these frameworks directly for analyzing the quality of different video explanations, however, falls short for at least two reasons. First, most frameworks summarize criteria based on a single research strand, e.g., on instructional explanations (e.g., Kulgemeyer, 2018a) or multimedia design (e.g., Brame, 2016), however, do not connect different research traditions. Thus, the focus is either only on content (instructional explanations) or only on the design of the videos (multimedia) although both are relevant and more importantly – interdependent factors of video quality. As a result, the focus lies on certain aspects of the video explanation in isolation rather than taking a more holistic view by including different aspects of quality. Second, some frameworks focus on a specific video type and/or a specific discipline (e.g., worked examples in math) and, consequently, include criteria that are not relevant when evaluating a video with different content in another discipline (e.g., the principle of labor division in economics).

Accordingly, the overall aim of this paper is to connect the different research traditions by identifying criteria that—when combined—can be used to evaluate the overall instructional quality of different video explanations. The second aim is to operationalize the criteria in a reliable measure, i.e., a rating framework, which can be used by researchers and practitioners to evaluate video explanations. To this end, in a first step, we identify relevant criteria in a theoretical framework by integrating existing theoretical and empirical literature regarding instructional explanations in economic education (Findeisen, 2017; Schopf & Zwischenbrugger, 2015; Schopf et al., 2019), video explanations (Kulgemeyer, 2018a) and findings from technology-enhanced learning research which highlight the effective use of multimedia principles to reduce cognitive load (e.g., Brame, 2016; Kay, 2014; Mayer et al., 2020). In a second step, we develop a rating framework and test its validity and reliability with 36 videos created by preservice economic teachers. For the generation of the theoretical framework, we mostly rely on reviews that compile results from different research disciplines and content areas. Therefore, the theoretical framework and the rating framework are to a large extent interdisciplinary. Although the rating framework was refined by using videos from future economic teachers, we, nevertheless, contribute a valid and reliable framework that can be used by researchers and practitioners of different domains to better assess the quality of video explanations.

As we draw mostly from the literature regarding instructional explanations in the classroom and multimedia and design principles, we first summarize the literature in those two fields in a more general manner. As our research aim is to identify broad quality criteria for videos that explain a certain principle, rather than demonstrate a certain behavior, we leave aside specific video features that are mainly analyzed in the context of observational learning and modeling examples, such as instructor characteristics (e.g., van Gog et al., 2014) or video perspective (e.g., Fiorella et al., 2017). Due to our goal to assess the quality of the video itself rather than its usage within a more complex learning setting, we did not dive deeply into the literature on interactive elements in videos (e.g., Delen et al., 2014) or combining videos with generative learning tasks (e.g., Fiorella et al., 2020).

After reviewing the literature and contrasting existing measurements for video explanations, we explain how our framework was developed based on previous work before

describing the framework in detail. In the following section, we describe the context of the videos which were used to define the rating framework and provide an overview of the psychometric quality of the instrument and the results.

2 Theoretical Background and Literature Review

2.1 Quality Criteria for Instructional Explanations

In different strands of research, criteria for the quality of instructional explanations have been developed and discussed (Findeisen, 2017; Kulgemeyer, 2018b; Lee & Anderson, 2013; Leinhardt, 2001; Schopf & Zwischenbrugger, 2015; Schopf et al., 2019; Wittwer & Renkl, 2008). Findeisen (2017) developed an explanation quality framework for economic and business education, which we use in the current study because it allows us to integrate the different research strands and to organize the previous research on the quality of instructional explanations. Accordingly, we structured the state of research regarding quality criteria along the categories of content, learner orientation, representation/design, language, and process structure.

The *content* of instructional explanations should be correct, accurate, and complete. In economic education, Schopf et al. (2019) argued that a learner should be able to understand *what* the content of the explanation—for example, a concept or a principle—is about, *why* it works, *how* it works and what it is useful for within the domain. In this regard, the relevance of the explained content for the domain should be clear for the learner by, for instance, exemplifying in which context a certain topic or principle is needed. Besides the correct use of technical terms, it is relevant to explain a subject topic stepwise and to give reasonable explanations about why certain steps are necessary and how they relate to the domain principle (Wittwer & Renkl, 2008).

The second category, *learner orientation*, means that an explanation should be adapted to the learner group. Prior knowledge is one of the most relevant factors in this regard (Wittwer & Renkl, 2008). Kulgemeyer (2018b) called this criterion “adaptation to the explainee” (p. 120). It is important to consider the learners since explanations can easily be too difficult or too easy. In addition to the learners’ prior knowledge, the content of an explanation should also be connected to other learner characteristics by, for example, taking their perspective on a certain topic into account and connecting the topic to their (daily) life experiences (Schopf et al., 2019). Lastly, explanations can be further adapted through interactions between explainer and learner when, for example, an explainer uses questions to assess the learner’s understanding and adjusts an explanation to the learner’s needs (Sevian & Gonsalves, 2008).

Third, the use of different *representations*, such as examples, visualizations, analogies, and models, is often included in criteria for effective explanations (e.g., Findeisen, 2017; Geelan, 2013; Kulgemeyer, 2018a; Schopf et al., 2019). Especially, visual representations such as graphs, diagrams, and charts are regularly used in explanations (for economic education, cf. Ring & Brahm, 2020; Vazquez & Chiang, 2014) as they provide structure and can guide learners in the construction of their own internal representations. Examples can support understanding as they promote the connection of domain knowledge to the learners’ everyday experiences. When examples are used, they must clearly represent the principle, should cover all aspects of the topic, and should ideally be taken from the student’s everyday life (Schopf et al., 2019).

The fourth category, *language*, pertains to an appropriate complexity level as explainers need to translate between domain language and everyday terms (Kulgemeyer & Schecker, 2013). Furthermore, avoiding vagueness as well as using body language and gestures might further add to the quality of the explanation (Brown, 2006).

Finally, a clear and coherent *process structure* helps learners to follow explanations. A short introduction—clarifying the topic or question—might set expectations and activate prior knowledge (e.g., Charalambous et al., 2011). A summary at the end as well as coherent argumentation in between reduce strain on the cognitive capacities of learners and might help them in their understanding.

2.2 Multimedia Design for Video Explanations

Video explanations can be seen as multimedia material since they combine spoken words and visual representations (Mayer, 2014). Before presenting specific frameworks in the next section, it is important to illustrate their theoretical basis. Most fundamentally, Mayer's (2014) cognitive theory of multimedia learning describes the process of learning with text and pictures with the central conclusion that the integration of the information in the visualization and the spoken text is a prerequisite for successful learning. The cognitive load framework assumes that learning is associated with different kinds of cognitive load, which are influenced by the inherent complexity of the task as well as the design of the learning material (Sweller, 2020). In combination, these two theories result in a very general recommendation: multimedia material leads to higher learning outcomes when it is designed in a way that helps learners to integrate audio and visual elements and at the same time, to reduce unnecessary cognitive load.

More specifically, researchers have identified multiple design principles that can be used as guidelines for the development of multimedia learning material (overview see Mayer, 2014). We will now focus on the most important design guidelines for our research goal, which are principles concerning the relationship of visual and audio elements in video explanations.

First, according to the signaling principle, multimedia signals can help guide learners' attention toward the most relevant information and to connect different modalities (van Gog, 2014). In a video explanation, one possible example of this principle is highlighting the part of the visualization that is currently explained. Second, rather than presenting the same information in different modalities, visual and audio should complement each other in a meaningful way (Low & Sweller, 2014). Video creators violate this so-called *modality-principle*, for example, when the visual element includes complete sentences, and the explanation encompasses reading these sentences. Third, according to the *temporal contiguity principle* (Mayer & Fiorella, 2014), it is important to present visual and spoken textual information at the same time, rather than first talking about a visual and then showing the visual (or the other way around). Fourth, video creators should refrain from using visuals with irrelevant (but possibly interesting) details, as they might distract the learner from the important content (Mayer & Fiorella, 2014).

After these foundational multimedia principles, we will now review different frameworks and measurements for video explanations.

2.3 Frameworks and Measurements for the Quality of Video Explanations

Multiple frameworks and measurements (see Table 1) describe and analyze the quality of video explanations, which are mainly influenced by two research traditions: multimedia and cognitive load research (e.g., Brame, 2016; Kay, 2014) as well as instructional explanations (e.g., Kulgemeyer, 2018a).

The different studies in Table 1 can be divided into two groups in terms of their objective. We use the term “guideline” to refer to studies that aim at providing recommendations to support instructors in developing and selecting more effective videos (Brame, 2016; Kay, 2014; Kulgemeyer, 2018a; Schopf, 2020; Siegel & Hensch, 2021). In comparison, we categorize studies as “measures” when they develop measurements for existing videos by operationalizing criteria in a coding manual (Kay & Ruttenberg-Rozen, 2020; Kulgemeyer & Peters, 2016; Marquardt, 2016). As we have already discussed the underlying research of what constitutes the quality of a video explanation by analyzing instructional explanations and multimedia and cognitive load research, we do not go into detail regarding the design recommendations. Instead, we shift our focus towards measures—and, thus, the question of how the quality of video explanations has been assessed so far.

Based on earlier frameworks of “traditional” explanations, Kulgemeyer and Peters (2016) analyzed the quality of video explanations for physics on YouTube. To measure the quality of the videos, they applied a dichotomous approach and rewarded each video with one point if a certain criterion was met (or subtracted one point for “negative” criteria, such as scientific mistakes) and used the total number of points as a measure of quality. Although they obtained high inter-rater reliability (Cohen’s kappa of $\kappa=0.860$) and satisfying internal consistency (Cronbach’s alpha of $\alpha=0.69$), the dichotomous nature of the instrument is focused on the occurrence of certain events (such as whether an equation is used to explain the content) and the subsequent use of the same category is irrelevant. This has two implications: first, it is difficult to transfer this measure, to a context where some category might not be relevant (e.g., where there is no equation). Second, it is not possible to identify more subtle qualitative differences between videos that meet the same criteria. For instance, the video quality differs most likely not only because the video does (or does not) use visualizations but because of the kind of visualization used and how it is connected to the (verbally explained) content (Mayer & Fiorella, 2014).

In the context of mathematics, Marquardt (2016) developed a rating scheme for video explanations with 22 criteria in four categories (overview in Table 1). The author operationalizes most criteria on a five-level scale and thus, theoretically, the rating scheme might be able to identify the differences between videos regarding the same criterion. Although the resulting measure combines multiple theoretical approaches, the authors did not test the rating scheme with videos and thus do not report consistency or reliability.

Also for mathematics, Kay and Ruttenberg-Rozen (2020) had students in teacher education generate video-based worked examples. Based on Kay’s (2014) framework, the quality of the student-generated video explanations was rated regarding four categories: establishing context ($n=3$ items), creating effective explanations ($n=7$ items), minimizing cognitive load ($n=4$ items), and engagement ($n=5$ items). All items were assessed on a three-point scale. The authors report acceptable internal consistency for each category-scale (Cronbach’s alpha between 0.60 and 0.85), but no inter-rater reliability. Due to the focus on worked examples, some aspects of quality, such as technical correctness or adaption to prior knowledge, are not part of the instrument.

Table 1 Overview of studies providing recommendations or measures for the quality of video explanations

Authors [Language of work]	Type/Aim	How are criteria identified?	Design recommendations/analyzed criteria	Domain/context
Brame (2016) [English]	<i>Guideline</i> Aim: identify design principles that allow instructors to choose or develop effective videos	Criteria derived from previous theoretical and empirical research in: Cognitive load & CTML Massive open online courses	A total of 13 criteria in three areas: Reducing cognitive load (e.g., by using signals to highlight important information) Student engagement (e.g., keep videos short) Active engagement (e.g., use guiding questions)	Biology
Kay (2014) [English]	<i>Guideline</i> Aim: identify design principles that allow instructors to choose or develop effective videos	Criteria derived from previous theoretical and empirical research in: Worked examples Instructional explanations Cognitive load	A total of 16 criteria in four areas: Establishing context (e.g., problem description) Effective explanations (e.g., stepwise solutions) Minimizing cognitive load (e.g., highlighting) Engagement (e.g., engaging voice)	Math
Kay and Rutenberg-Rozen (2020)* [English]	<i>Measure</i> Aim: develop and test a measure to determine the quality of existing videos	See Kay (2014)	See Kay (2014)	Math
Kulgemeyer (2018a) [English]	<i>Guideline</i> Aim: identify principles that allow instructors to choose or develop effective videos	Criteria first derived from previous theoretical and empirical research in: Instructional explanations Videos based on criteria are then tested in experimental design	A total of 14 criteria in seven areas: Structure (e.g., summary) Adaptation (e.g., adaptation to prior knowledge) Tools for adaptations (e.g., examples) Minimal explanation (e.g., high coherence) Highlighting relevancy (e.g., direct addressing) Follow-up learning task New complex principle	Physics

Table 1 (continued)

Authors [Language of work]	Type/Aim	How are criteria identified?	Design recommendations/analyzed criteria	Domain/context
Kulgemeyer and Peters (2016) [English]	<i>Measure</i> Aim: develop and test a measure to determine the quality of existing videos	Criteria derived from previous theoretical and empirical research in: Instructional explanations	A total of 31 criteria in seven areas: Content (e.g., technical correctness) Structure (e.g., giving an outlook) Use of language (e.g., paraphrasing technical terms) Context and examples (e.g., addressing explainee) Mathematics (e.g., using formulae) Interrogation (asking further questions) Non-verbal elements (e.g., using visualizations)	Physics
Marquardt (2016) [German]	<i>Measure</i> Aim: develop and test a measure to determine the quality of existing videos	Criteria derived from previous theoretical and empirical research in: Instructional explanations Cognitive load & CTML Media pedagogy/quality of learning material	A total of 22 criteria in four areas: Didactics/content (e.g., technical correctness) Didactics/methods (e.g., adaptation to the learner) Media design (e.g., highlighting the most relevant information) Pedagogical aspects (e.g., discussion of learning goals)	Math

Table 1 (continued)

Authors [Language of work]	Type/Aim	How are criteria identified?	Design recommendations/analyzed criteria	Domain/context
Schopf (2020)* [German]	<i>Guideline</i> Aim: identify principles that allow instructors to choose or develop effective videos	Criteria derived from previous theoretical and empirical research in: Instructional explanations Cognitive load & CTML	An undefined number of criteria, in three main areas: Video context features (e.g., short video length and interactivity) Reducing cognitive load (e.g., signaling) Content and learner orientation (e.g., adaptation to prior knowledge, technical correctness and completeness, examples)	Economics
Siegel and Hensch (2021)* [German]	<i>Guideline</i> Aim: identify principles that allow instructors to choose or develop effective videos	Criteria derived from previous theoretical and empirical research in: Instructional explanations Cognitive load & CTML Media pedagogy/quality of learning material	A total of 20 criteria in 5 areas: Didactics (e.g., learner orientation) Psychology (e.g., reducing cognitive load) Film (e.g., narration) Technology (e.g., image quality) Legal (e.g., copyright)	Multiple

* As our theoretical framework was developed in 2020, these resources were not known to us at the time of instrument development but are included to mirror the current state of research

Table 2 Framework for the quality of video explanations: overview

Category	Criteria	Definition	Originally derived from	Similar criteria in
Content	Technical correctness	The video explanation is technically correct if the mentioned relationships, technical terms, and contents are explained correctly and precisely	Correct use of technical terms (Find-eisen, 2017) Technically correct (Schopf & Zwis-chenbrunner, 2015)	Kulgemeyer and Peters (2016) Marquardt (2016) Schopf (2020)* Siegel and Hensch (2021)*
	Technical completeness	The video explanation is techni-cally complete if no information or subject-specific terms relevant to the topic or the argument are omitted	Technical completeness (Find-eisen, 2017) Core statements used to answer the questions What/how/why (Schopf & Zwischenbrunner, 2015)	Kay (2014)/ Kay and Ruittenberg-Rozen (2020)* Marquardt (2016) Schopf (2020)* Siegel and Hensch (2021)*
Learner orientation	Relevance to the learners	The video explanation establishes relevance to the learners when their perspective is considered and thus the content is aligned with the learners' experiences	Linking to entry requirements and connecting with prior content (Schopf & Zwischenbrunner, 2015) consideration of learners' prior knowledge and characteristics (Find-eisen, 2017) Highlighting relevance (Kulgemeyer, 2018a)	Kay (2014)/ Kay and Ruittenberg-Rozen (2020)* Marquardt (2016) Kulgemeyer and Peters (2016) Marquardt (2016) Schopf (2020)* Siegel and Hensch (2021)*
Learner orientation	Linking to prior knowledge	The video explanation links mean-ingly to prior knowledge if it is adapted to the learners and their prior subject knowledge in terms of the complexity and scope of the content	linking to entry requirements and connecting with prior content (Schopf & Zwischenbrunner, 2015) consideration of learners' prior knowledge and characteristics (Find-eisen, 2017) adaption (Kulgemeyer, 2018a)	Marquardt (2016) Schopf (2020)* Siegel and Hensch (2021)*
Learner orientation	Active engagement	Learners are encouraged to actively participate if the explainer video contains tasks that learners must complete during the video	Not derived theoretically (see appen-dix 5 for more details)	Brame (2016) Kay (2014)/ Kay and Ruittenberg-Rozen (2020)* Siegel and Hensch (2021)*

Table 2 (continued)

Category	Criteria	Definition	Originally derived from	Similar criteria in
Learner orientation	Direct addressing	Learners are addressed directly in the video explanation when the speaker uses second-person singular or first-person plural	Direct addressing (Kulgemeyer, 2018a)	Brame (2016) Kulgemeyer and Peters (2016) Marquardt (2016) Siegel and Hensch (2021)*
Representation and design	Use of examples	Examples are used in the video explanation to promote learning if they clarify the domain content and represent the underlying principle	Examples (Findeisen, 2017; Kulgemeyer, 2018a, 2018b; Schopf & Zwischenbrugger, 2015)	Kulgemeyer and Peters (2016) Marquardt (2016) Schopf (2020)* Siegel and Hensch (2021)*
Representation and design	Design of visualizations	Visualizations are used in the video explanation to promote learning if they are technically correct and clearly designed	Principles are visualized adequately (Findeisen, 2017; Schopf & Zwischenbrugger, 2015) Representation forms and demonstration (Kulgemeyer, 2018a) Multimedia research (coherence, contiguity, signaling e.g., in Mayer, 2014)	Brame (2016) Kay (2014)/Kay and Ruitenberg-Rozen (2020)* Kulgemeyer and Peters (2016) Marquardt (2016) Schopf (2020)* Siegel and Hensch (2021)*
Representation and design	Matching of visualizations and spoken text	Visualizations and spoken content are used in the video explanation to promote learning when they are coordinated in terms of content and timing	Principles are visualized adequately (Findeisen, 2017; Schopf & Zwischenbrugger, 2015) Different representational formats and their integration (Findeisen, 2017) Multimedia research (modality and redundancy in Mayer, 2014)	Brame (2016) Kay (2014)/Kay and Ruitenberg-Rozen (2020)* Kulgemeyer and Peters (2016) Marquardt (2016) Schopf (2020)* Siegel and Hensch (2021)*
Language	Comprehensive language	The language of the video explanation is comprehensive if it is kept as simple as possible in terms of grammar and use of foreign words	Simple language, Level of language (Kulgemeyer, 2018a; Schopf & Zwischenbrugger, 2015)	Brame (2016) Marquardt (2016) Schopf (2020)* Siegel and Hensch (2021)*

Table 2 (continued)

Category	Criteria	Definition	Originally derived from	Similar criteria in
Language	Precise language	The language of the video explanation is precise when words are pronounced clearly and correctly, and speech tempo and pauses are used appropriately	Precise language (Findeisen, 2017)	Brame (2016) Marquardt (2016) Schopf (2020)* Siegel and Hensch (2021)*
Process structure	Structure	The video explanation is well structured when the argumentation is coherent	Emphasizing the objective of the explanation (Findeisen, 2017) Summary (Findeisen, 2017; Kulgemeyer, 2018a) Structure (Schopf & Zwischenbrugger, 2015)	Kay (2014)/Kay and Ruitenberg-Rozen (2020)* Kulgemeyer and Peters (2016) Marquardt (2016) Siegel and Hensch (2021)*
Only included in the theoretical framework, not in the rating framework				
Process structure	Context	The video is well integrated into the context when it is embedded in ongoing learning activities and combined with a follow-up learning task	Follow-up Learning Task (Findeisen, 2017; Kulgemeyer, 2018a; Schopf & Zwischenbrugger, 2015)	Brame (2016) Kay (2014)/Kay and Ruitenberg-Rozen (2020)* Siegel and Hensch (2021)*

* As our theoretical framework was developed in 2020, these resources were not known to us at the time of instrument development but are included to mirror the current state of research

In summary, previous instruments have different limitations, which are the reason for the development of a new measure: First, from a theoretical point of view, they do not include all relevant criteria or include criteria that are not easily transferable to other contexts (e.g., Kay & Ruttenberg-Rozen, 2020). Second, from a methodical perspective, they lack evidence regarding interrater reliability and internal consistency (e.g., Marquardt, 2016) or due to a dichotomous approach do not provide enough information about variance within a criterion (Kulgemeyer & Peters, 2016).

3 Development of the Rating Framework

3.1 Development Process

The rating framework was developed in two steps: Before the video rating, we theoretically derived an initial framework based on research on instructional explanations in economic education (Findeisen, 2017; Schopf & Zwischenbrugger, 2015), multimedia design principles (overview in Mayer, 2014) and video explanations (Kulgemeyer, 2018a). These resources were not chosen based on a systematic search but because they combine different research approaches and thus, taken together, provide valid criteria for video explanations as each offers a unique and relevant perspective: Findeisen (2017) and Schopf and Zwischenbrugger (2015) were relevant because of their focus on the content of explanations in economic education, the design principles in Mayer (2014) explain the effect of multimedia learning in a more general manner (see Sect. 2.2) and Kulgemeyer (2018a) made use of instructional explanation literature to analyze the quality of video explanations. From these resources, we identified relevant criteria and structured them according to the five categories that Findeisen (2017) already used to describe the quality of instructional explanations in the classroom: content, learner orientation, representation/ design, language, and process structure (see Sect 2.1). Although the categories were developed for explanations in the classroom, we used them because they fit the overall aim to develop a comprehensive framework and presented categories that could be used to subsume criteria from the chosen literature. We then compared the results with some of the frameworks and instruments that are described in Sect. 2.3 to check whether we reached similar criteria (see Table 2).

In the second step, five of the videos described in Sect. 4.1 were chosen at random to test and inductively revise the rating framework to reach a usable coding manual. With this, we strived to ensure that (a) most criteria were developed before the viewing of the material and that (b) it was still possible to change, add, or omit criteria based on the actual material. This procedure as well as the theoretical framework itself were preregistered to make the changes resulting from the second phase more transparent and verifiable.¹

With the resulting coding manual, all videos were then rated by two raters. The raters were the first author of this paper and a research assistant. The five videos that were used to develop the coding manual served as anchor examples. For all criteria, the raters documented not only their final rating but also their reasoning in case their assessment differed from the highest possible rating (see Appendix 4 for the rating sheet template). After the first ratings, deviating from our preregistered analysis plan, we sent the coding manual to

¹ The preregistration, including the theoretical framework for this research project, can be viewed and downloaded here: https://osf.io/gepbu/?view_only=47c1cab6cee14dca8580b65bab6a6f41. The data, code and framework can be found here: doi.org/10.17605/OSF.IO/AKBWJ.

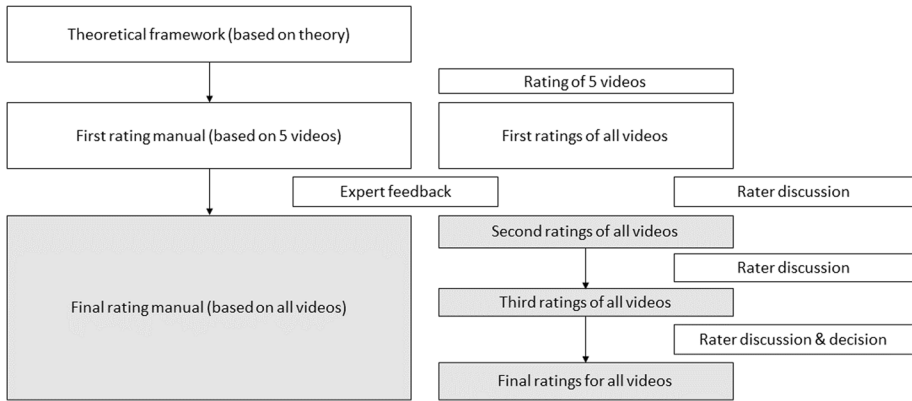


Fig. 1 Overview of the development procedure of the rating framework. *Note* The grey areas show the framework (and the resulting data) that are reported in this manuscript

one expert for video ratings as well as one expert for instructional explanations in economic education and asked for feedback regarding validity and comprehensibility. The expert feedback was added to our research because the first ratings resulted in low-inter-rater reliability for some criteria. Based on the feedback, we adjusted the manual again.

In total, the videos were rated three times. An overview of the whole procedure can be found in Fig. 1, which also shows that the rating framework remained the same from the second rating of all videos. The final coding manual as described in Sect. 3.2 differs from the theoretical framework (that was developed before analyzing the data) in several regards which are made transparent in Appendix 5.

3.2 Rating Framework for Video Explanations

After the final adjustments of the theoretical framework based on the five preselected videos, the coding manual consisted of twelve criteria in five categories (see Table 2 for an overview): (1) content, (2) learner orientation, (3) representation and design, (4) language, and (5) process structure. All criteria were rated on a scale with four levels to assess to what extent each criterion was fulfilled (0 = not or only barely fulfilled, 1 = partly fulfilled, 2 = mostly fulfilled, 3 = always fulfilled). For instance, when a video included a mistake at the beginning but was otherwise flawless, we rated technical correctness as mostly fulfilled (=2). A rating of 0 was only assigned when the criterion was not fulfilled throughout the whole video. For all criteria, a definition and the relevant conditions were outlined in bullet points in the coding manual. The complete coding manual can be found in Appendix 2.

Regarding content, the first criterion was **technical correctness**, i.e., no errors in the explained content. Videos received a lower rating when they included technical errors or imprecise statements. One video, for instance, described the stock market and used an example of a very small company to illustrate its stock market launch. In the description of the initial public offering, the company sold seven shares for 1,000€ each. This, however, would not be possible in Germany as a share capital of at least 50,000€ is necessary for a company to go public. As this is clearly a technical error, technical correctness was seen as only mostly, not completely fulfilled.

The second criterion, **technical completeness**, was achieved when no relevant information or subject-specific terms were missing. As all videos had different content, expectations were not predefined instead the raters assessed the relevant information for the topic while rating the video based on their expertise and the economics curriculum. Although a video can't encompass all information regarding a certain topic, the most important information needed to be included. One video, for instance, explained how the Gross Domestic Product (GDP) is calculated and used as a measure of economic growth. Typical limitations and criticism of the GDP, however, were not part of the video although this was expected by the raters based on the curriculum.

In the category learner orientation, the first criterion was **relevance to the learners**. A high rating was achieved when the learner's perspective was considered in the video by, for example, connecting the content to the learner's everyday life or introducing a fictitious character that might represent the learner's perspective. In a video that explained the effect of taxes, for instance, the topic was not connected to the learner's experiences with taxes—which could have been done, for example, by starting with the question of where and how the learner might pay taxes. Instead, the video explained gasoline taxes as an example although the targeted learners are probably not familiar with gas taxes as they are not legally allowed to drive.

Regarding the criterion **linking to prior knowledge**, the raters evaluated to what extent the content was related to the learner's prior knowledge in terms of complexity and scope. Again, prior knowledge was not predefined but was instead assessed by the raters based on the grade level indicated for the video explanation and the curriculum. Videos would receive lower ratings when new subject-specific terms or principles were introduced without an explanation or already known principles were discussed in detail.

The third criterion in this category, **active engagement**, was fulfilled when learners were given a task that might lead to active participation; for instance, they were asked to pause the video to consider examples or assess their understanding. Not only the number of such integrated tasks but also their quality was used to form a rating.

Direct addressing was met when the learner was directly addressed by the speaker, e.g., by using the second-person singular or first-person plural instead of the passive voice. This was only consistently used in a few videos. For example, in a video about the calculation of the GDP, the speaker started by connecting the content to prior knowledge, "as you already know, the GDP...", and used the first-person plural when explaining the new content, "we should not forget, however, that intermediate inputs must be deducted before...".

In the category design and representation, the criterion **use of examples** was fulfilled when the video used appropriate, comprehensive, and authentic examples to illustrate the content. Note that whether examples were a good fit regarding the learner's everyday experiences was rated in the above criterion relevance to the learners. The number of as well as the quality of examples were considered in the rating. A suboptimal rating was often due to the use of very general instead of specific examples. For example, in a video where the labor market was explained, an example illustrated how employers and employees discuss certain wages. Instead of using a specific company, a specific employee, and a specific wage, the video remained abstract.

All aspects of visualizations that could be assessed without considering the audio track were rated in the **design of visualizations**. Here, different aspects of the visualizations were considered together to form a rating. The highest rating was achieved if the visualizations were error-free and when the video was not overloaded with too many visualizations (or barely contained any visualizations). For more complex visualizations, missing step-by-step construction and no signals also led to a lower rating. As such, a video with too many

A: Low rating for design of visualizations



B: Low rating for matching of visualizations

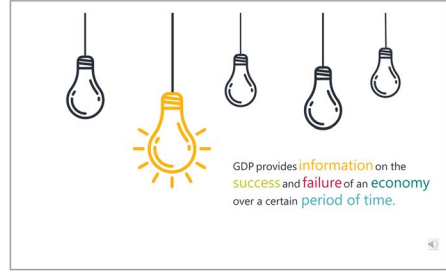


Fig. 2 Screenshots of two videos with low ratings regarding design (A, left panel) or matching of visualizations and spoken text (B, right panel)

visualizations received a low rating (example in Fig. 2, left panel), and a video with only one complex visualization also received a lower rating, when no signals and/or no step-by-step construction were used to guide the learner's attention.

For the criterion **matching of spoken text and visualizations**, the raters assessed whether spoken text and visualizations were linked in a way that promoted learning. For this criterion, the videos received a lower rating when general visualizations were used that did not match the specific video content (example in Fig. 2, right panel) or when temporal contiguity was not met. Other conditions in this criterion were coherence (no unnecessary/seductive details in visualization) and redundancy (only keywords of spoken text were allowed as written text).

Regarding language, the criterion **comprehensive language** was fulfilled if the language syntax and word usage were kept as understandable as possible. The videos received lower ratings when the speaker used unnecessary foreign or complex words or when their sentences had long or multiple subordinate clauses.

The video received a high rating for **precise language** when the speaker's voice was accent- and mostly dialect-free, when the sentences were complete and free of errors, and when appropriate pauses were used. Videos with computer-generated voices, for example, received a lower rating as they were characterized by unusual intonation and unclear pauses.

In terms of process structure, videos received the highest rating regarding **structure** when the objective, topic, or question was clearly defined at the beginning of the video, when there was a coherent argumentation structure, and when there was a clear ending with a summary, follow-up task, or transition to a new topic.

4 Applying the Rating Manual with Videos from Preservice Teachers

4.1 Sample: Videos and Their Creators

For this study, we used videos that were created by $N=36$ preservice economics teachers (16 females, $M_{\text{age}}=24.71$ years, $SD=3.00$). For one semester (April–July 2020), the students participated in two different courses that are part of the curriculum for preservice economic teachers. As part of the seminar, they were asked to design a video explanation. Their task was to choose a tool in which spoken text and visualizations could be combined

Table 3 Inter-rater reliability for the video quality ratings

Category	Criteria	Second ratings		Third ratings	
		Fleiss' Kappa	ICC	Fleiss' kappa	ICC
Content	Technical correctness (C1)	0.44	0.79	0.86	0.96
Content	Technical completeness (C2)	0.28	0.64	0.82	0.93
Learner orientation	Relevance to the learners (L1)	0.74	0.94	0.92	0.99
Learner orientation	Linking to prior knowledge (L2)	0.13	0.56	0.75	0.9
Learner orientation	Active engagement (L3)	0.63	0.86	0.92	0.99
Learner orientation	Direct addressing (L4)	0.52	0.82	0.84	0.97
Representation and design	Use of examples (R1)	0.43	0.87	0.89	0.98
Representation and design	Design of visualizations (R2)	0.46	0.87	0.75	0.94
Representation and design	Matching of visualizations and spoken text (R3)	0.37	0.72	0.7	0.91
Language	Comprehensive language (S1)	0.46	0.82	0.61	0.87
Language	Precise language (S2)	0.51	0.91	0.79	0.96
Process structure	Structure (P1)	0.42	0.75	0.9	0.97

to create a three- to seven-minute-long video. Most videos were created with Simpleshow (<http://simpleshow.com/>) or Presentation slides with voice input. Some students also used Adobe Spark (adobe.com) or recorded their screens (while working with a visualization app). The video rating made up part of the grade for the seminar to motivate students to create a high-quality video that they might later use as part of their teaching. Before the video assignment, the participants received information regarding the relevance and quality of instructional explanations (based on Schopf et al., 2019) but no information regarding high-quality video explanations or multimedia design principles. They received a short introduction to the tools and were given three weeks to create the video. To make sure that the videos had different topics and, thus, would be an appropriate sample to test the applicability of the framework, the students could choose the topic out of a list based on the economics curriculum. In total, 36 videos were used to validate the framework for video explanations in economic education. A list of all videos, including topic and length, can be found in Appendix 1.

4.2 Results

4.2.1 Reliability

In line with the preregistration, Fleiss' kappa was used to determine the inter-rater reliability for all categories separately after the second and third ratings (see Table 3). Additionally, we report a two-way mixed, agreement, average-measures intraclass correlation (ICC) as this is more suited to the ordinal nature of our variables (Hallgren, 2012).

After the second rating (i.e., the first rating with the final manual), we did not reach high interrater reliability for all categories. Especially for ratings regarding technical completeness as well as adapting to prior knowledge, the agreement was rather low. As we did not predefine aspects that should be included or prior knowledge that could be used as a basis for the two ratings respectively, we found these criteria difficult to assess objectively. Since we could not attribute our differences to an unclear understanding of the rating framework,

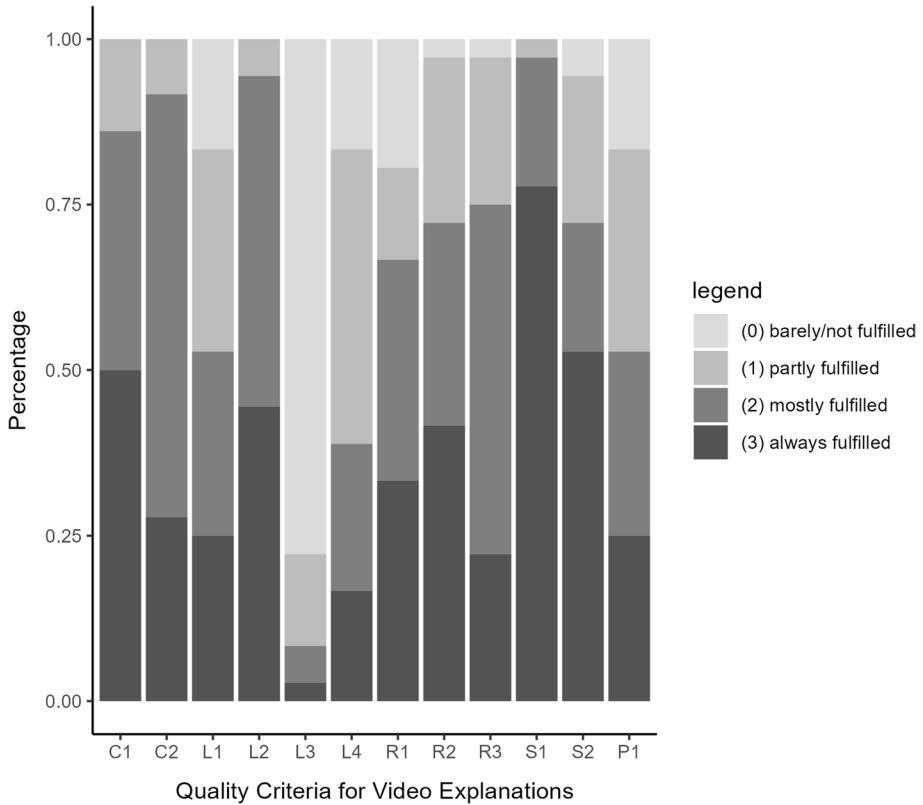


Fig. 3 Overview of the relative frequency of the different ratings for all criteria

we decided not to revise the framework but rather discuss our understanding of all criteria and rate the videos a third and final time, which led to an increase in reliability. We discuss reliability as a limitation in more detail in the last section of the paper. After the third rating, the reliability for most criteria was considered good (Fleiss' kappa > 0.61) or almost perfect (Fleiss' kappa > 0.81, Landis & Koch, 1977). Based on intraclass-correlation, the agreement was already excellent for most criteria (ICC > 0.75 according to Cicchetti, 1994) for the second rating.

All disputes after the third rating were resolved by discussions among the raters. Not only the ratings but also the reasoning behind the ratings were used as a basis for the final decision. The final ratings were used to assess internal consistency for the complete scale. Cronbach's alphas for the scale consisting of all 12 criteria was 0.73. Due to the small sample size and the small number of criteria for some categories, factor analysis does not yield useful results. An overview of the relationship between the different ratings in a correlation matrix can be found in Appendix 3.

4.2.2 Overview of the Ratings

Figure 3 shows an overview of the range of the quality ratings by visualizing the relative frequency of the four levels for each criterion. For some criteria, such as technical

correctness (C1), the lowest rating was not used at all for the videos in this study and half of the videos received the highest rating. Active engagement (L3) was rated as barely or not fulfilled for almost all videos. For process structure (P1), all possible ratings were used in a similar frequency.

5 Discussion and Conclusion

The aim of this study was twofold: First, we identified relevant criteria from the literature to develop a theoretical framework based on existing theoretical and empirical work on instructional explanations in economic education (Findeisen, 2017; Schopf & Zwischenbrugger, 2015), video explanations (Kulgemeyer, 2018a), and multimedia design principles (overview in Mayer, 2014). We thus contribute a theory-driven instrument to assess the quality of video explanations in economic education to the literature which, due to its rather broad definition of criteria might be used more generally in different domains. Second, we investigated the psychometric quality of a coding manual with twelve criteria based on 36 videos that had been created by preservice teachers in a university seminar. We found the coding manual to be (mostly) reliable and the range of quality ratings fit the context of our videos. The results—both the instrument and its application—contribute to the current literature in several ways and have implications for future research and practice.

The instrument can be used in future research as well as practice to evaluate the quality of existing video explanations. It goes beyond earlier frameworks by integrating conceptual and empirical research from multimedia research and research on instructional explanations in different domains (Findeisen, 2017; Kulgemeyer, 2018b; Lee & Anderson, 2013; Leinhardt, 2001; Schopf & Zwischenbrugger, 2015; Schopf et al., 2019; Wittwer & Renkl, 2008). All criteria which we had identified as relevant and included in our instrument, have in the meantime (i.e., since the development of the instrument) also been part of newer frameworks dealing with video explanations (Kay & Ruttenberg-Rozen, 2020; Schopf, 2020; Siegel & Hensch, 2021) which can also be taken as an indicator for the instrument's content validity. In line with similar frameworks and recommendations (Brame, 2016; Kay, 2014; Schopf, 2020; Siegel & Hensch, 2021), the instrument can also be used as a guideline for the creation of new video explanations by, for example, teachers, teacher educators, or students. One advantage of the framework presented here is that it is not thematically limited to a certain topic or a certain context (such as worked examples) and thus provides a flexible framework.

Before the framework and the coding manual are used in future research, however, it is important to discuss the potential boundaries of the instrument. The overall focus of the instrument is on the content of the videos. Thus, a clear limitation is that other relevant conditions, such as how the video is embedded in the greater learning context or how it should be adjusted to different learners, were not considered.

Nevertheless, the instrument is valuable because it enables teachers to systematically evaluate stand-alone videos which are now widely available on various platforms. For video explanations that are used in combination with other material or as part of a certain educational setting (such as a flipped classroom), the framework could be altered in future research to include aspects that are relevant to the respective setting.

Since the videos had different topics, we decided not to define the expectations regarding content and learners' prior knowledge before the rating. Although this results in broader applicability of the framework, at the same time, it increases the subjectivity of

the assessment as the raters need to evaluate these criteria based on the curriculum and their expertise. This, however, might lead to a higher need for rater coordination—which was also visible in our data as the inter-rater reliability of two criteria was very low after the first and second ratings and only increased after the last rater discussion. If multiple videos with the same content are evaluated with the framework, a clear definition of expectations regarding content and learners' prior knowledge would increase the objectivity of the instrument.

Furthermore, we often combined multiple conditions for a criterion to be met. For example, for *matching of visualization and text*, we rated not only temporal and spatial proximity but also coherence, consistency, and redundancy. Even though this makes it easier to compare the videos with regard to the use of visuals in general and leads to a higher variance concerning the criterion, it still means a loss of information compared to a separate rating for all of the conditions. For instance, when a video has a lower rating regarding the criterion, the rating alone cannot be used to identify which of the conditions were not fulfilled, i.e., whether the video lacked coherence or temporal proximity. Especially when concentrating on the role of visualizations and their relation to the spoken text, it seems advisable to further develop the criterion by splitting it into different criteria.

To better evaluate the results and validity of the instrument, we can compare our results to existing empirical research on the quality of instructional explanations in the classroom (Findeisen, 2017). In the category *content*, most videos received higher ratings whereas this category was challenging for preservice teachers in authentic explanation settings (Findeisen, 2017). One reason for this could be the difference between video explanations and instructional explanations: For the video explanations, the preservice teachers had the time and opportunity to check the technical correctness of the content and to repeat the “production” process if necessary. Thus, errors seem to be less likely for video explanations.

In the category *learner orientation*, our results are somewhat comparable to the findings for instructional explanations in the classroom. Although Findeisen stated that most preservice teachers were able to adapt the content to the learners, she also argued that the evaluation of prior knowledge is one of the major challenges (Findeisen, 2017). Besides adapting the content to the learners' prior knowledge, establishing the relevance of the topic by connecting it to the learners' everyday experiences was an additional challenge for the preservice teachers who created the videos. Actively involving, i.e., cognitively activating the learners through tasks, was not a priority for the preservice students in our sample. A potential explanation for this is that the preservice teachers do not see the need for such cognitive activation even though the literature highlights the need for this (Brame, 2016). Furthermore, actively engaging learners was not prompted by the tools used in this study. When other software is used—for example, H5P (h5p.org)—the share of videos with more interaction would probably be higher. Although directly addressing could further highlight the relevance of the content to the learners (Kulgemeyer, 2018a), it might also be a technique that the preservice teachers are not accustomed to from existing educational videos on the internet and, consequently, do not apply to their videos.

Regarding *representation and design*, visualizations are often seen as a more challenging aspect in instructional explanations (Findeisen, 2017; Schopf et al., 2019). We could partly replicate this finding for video explanations although visualizations were often used adequately while the videos were rated lower regarding the design and combination with spoken text. Fewer errors in visualizations could be explained because most visualizations used in the videos were not created by the preservice teachers while explaining (and thus were more likely a result of a systematic searching process and not a spontaneous byproduct of the explanation). Regarding the lower ratings for the combination of text and

visualizations, it could be argued that preservice teachers have little knowledge of how to design multimedia material in a way that promotes learning. Here, the deployed tools could also have some impact as they provide different opportunities or have certain default settings that might influence the combination of spoken text and visualizations. One tool, for example, uses certain spoken words as cues for the appearance of visualizations and thus automatically leads to high temporal contiguity. As a practical implication of our research, it seems helpful for future teachers to develop knowledge of multimedia principles as part of their teacher training because the development (or evaluation/augmentation) of learning material will most likely be part of their job in light of continuing digitalization.

While being precise was a challenge for the preservice teachers, comprehensive language did not pose a problem for the majority of future teachers. The lower ratings for precision can be partly explained by the fact that not all future teachers wanted to record their own voices and computer voices automatically received a lower rating.

In terms of process structure, most videos had a clear structure. A possible explanation for not including a clear ending might be that the preservice teachers deemed the videos to be too short to provide a summary at the end.

Up to now, the assessment of video quality has only been based on the videos themselves. Consequently, further analyses of the videos are necessary to determine whether the aspects that (theoretically) should influence the quality of the video explanation are indeed beneficial for the learner. Following Kulgemeyer (2018a), one possible approach might be to develop videos that systematically differ regarding the criteria and to test their effect on learners. It could be assumed that different criteria would affect the learners differently. Therefore, such a study might help to further develop the measure as the criteria could be weighted according to their empirical effects.

Appendix

Appendix 1: List of videos

Topic	Grade level	Length
Perfect competition	8	04:38
Oligopolies	8	04:32
Efficiency	8	07:38
Pricing	8	04:37
Monopolies	8	04:26
Sustainable management	8	02:43
The flow of income model (simple version)*	8	03:55
The flow of income model (advanced version)	8	04:11
Opportunity costs	8	03:18
Stock exchange*	8	06:31
Different types of goods	8	04:22
Supply and demand function	8	03:56
Determining factors of supply and demand	8	07:26
Corporate goals	9	06:26
Labor market	9	03:13

Topic	Grade level	Length
Influence of taxes	9	05:08
Operational procedures	9	05:23
GDP*	10	03:52
Influence of subsidies	10	03:01
Market regulation	10	04:45
Business Cycle*	10	06:04
Public goods	11/12	05:43
Social market economy	11/12	08:19
Protectionism	11/12	02:42
The role of the state in the flow of income model	11/12	02:55
Business cycle	11/12	05:00
Free trade	11/12	04:33
Keynesianism	11/12	06:57
GDP calculation	11/12	02:49
Neoliberalism	11/12	02:50
Alternative wealth indicators	11/12	05:50
Macroeconomic equilibrium	11/12	06:09
Economic policy	11/12	03:12
European Central Bank	11/12	04:18
Stock exchange	11/12	04:44
Freedom rights*	11/12	03:47

*videos were used in the development of the coding manual

Appendix 2: Coding Manual

Technical correctness (C1).

Definition: The video explanation is technically correct if the mentioned relationships, technical terms, and contents are explained correctly and precisely

Criteria: Type and quantity of errors
Type and quantity of imprecise statements

Ratings:

Rating	Description	Examples for given reasons	Example Video IDs
3	The video explanation is always correct	(No reason needed)	WU0103
2	The video explanation is mostly correct	A clear technical error: In the video, the impression is given that you can go public with 7,000 share capital, but you need at least 50,000 share capital	XB0012

Rating	Description	Examples for given reasons	Example Video IDs
1	The video explanation is partly correct	Several "small" technical errors or imprecise statements: "Our economic system is based on this theory," "Not everyone in society can afford a car, therefore not everyone has a car," "People who pay a lot of taxes have no influence on what the money is used for."	HH1628
0	The video explanation is barely correct, if at all	(No expression in the data)	–

Notes on demarcation: What is not part of this criterion?

List of aspects that do not fit this criterion:	Should be rated in:
One important subject area was completely omitted	Technical completeness (C2)
An axis label is missing from a graph	Design of visualizations (R2)
An example is not well suited to illustrate the principle	Examples (R1)
There is far too much content for learners with little prior knowledge	Prior knowledge (L2)

Technical completeness (C2).

Definition: The video explanation is technically complete if no information or subject-specific terms relevant to the topic or the argument are omitted

Two possibilities:

- Many videos on the same topic: Relevant technical terms or aspects of the topic are predefined in a document and the rating is done by comparing the content with the document
- Many videos on different topics: Relevant information and topic aspects are estimated by a domain expert based on the curriculum

Criteria: General criteria for (b):

- What/how/why questions are answered
- Why the content/model is relevant from a professional point of view is made clear
- Critical aspects are discussed or critical aspects are pointed out (controversial topics are discussed controversially)

Ratings:

Rating	Description	Examples for given reasons	Example Video IDs
3	The content of the video explanation consistently meets professional expectations with regard to the information, subject aspects, and technical terms mentioned	(No reason needed)	IY1506
2	The content of the video explanation mostly meets professional expectations with regard to the information, subject aspects, and technical terms mentioned	There is no justification for why goods have to be classified at all	XB0014

A Rating Framework for the Quality of Video Explanations

Rating	Description	Examples for given reasons	Example Video IDs
1	The content of the video explanation partly meets professional expectations with regard to the information, subject aspects, and technical terms mentioned	There is a lack of criticism of subsidies as well as different types of subsidies (Direct, Indirect, etc.)	XB0008
0	The content of the video explanation barely meets or does not meet professional expectations with regard to the information, subject aspects, and technical terms mentioned	(No expression in the data)	

Notes on demarcation: What is not part of this criterion?

List of aspects that do not fit this criterion:	Should be rated in:
Statement on aspect X would have to be more precise to be correct	Technical correctness (C1)
The video is overall clearly too easy for the grade level	Prior knowledge (L2)

Relevance to the learners (L1).

Definition: The video explanation establishes relevance to the learners when their perspective is considered and thus the content is aligned with the learners' experiences

Criteria: Examples are from the everyday lives of the learners
 Content is linked to the everyday lives of the learners (for example, through an "uninformed" character)
 Content is adapted to the age of the learners (e.g., not too childish or playful)

Ratings:

Rating	Description	Examples for given reasons	Example Video IDs
3	In the video explanation, the content is completely aligned with the learners' experiences	(No reason needed)	JO0710
2	In the video explanation, the content is mostly aligned with the learners' experiences	The examples are from the everyday lives of the learners (saving, receiving interest, child benefit, etc.), but the content could have been better and more consistently linked to the life worlds of the learners in some places	XB0009
1	In the video explanation, the content is partly aligned with the learners' experiences	Connection only partially given—for example, at the beginning—but then again, in many places, it is very general and not at all linked to the everyday world (would have been relatively easy with appropriate examples)	XB0014
0	In the video explanation, the content is not or only barely aligned with the learners' experiences	The perspective of the students is hardly taken; there is no linkage via examples from the everyday world or other points of reference	VN1306

Notes on demarcation: What is not part of this criterion?

List of aspects that do not fit this criterion: Should be rated in:

The example is from the everyday world but not typical for the principle Examples (R1)

Linking to prior knowledge (L2).

Definition: The video explanation links meaningfully to prior knowledge if it is adapted to the learners and their prior subject knowledge in terms of the complexity and scope of the content

Two possibilities:

(a) Many videos on the same topic: Relevant prior knowledge is defined in advance in a document and the rating is done by comparing the content with the document

(b) Many videos on different topics: Relevant prior knowledge is estimated by a domain expert based on the indicated grade level and the curriculum, and the rating is done by matching the individual estimate with the video content (the estimate is not pre-recorded)

Criteria: The amount of new content is adapted for the length of the video to the prior knowledge of the learners

The complexity of the content is adapted to the prior knowledge of the learners for the length of the video

Ratings:

Rating	Description	Examples for given reasons	Example Video IDs
3	The content of the video explanation is completely linked to prior knowledge	(No reason needed)	XB0005
2	The content of the video explanation is mostly linked to prior knowledge	Overall a bit too fast paced even for upper level	XM0005
1	The content of the video explanation is partly linked to prior knowledge	More than one term that would need to be explained for the 8th grade: transactions, sectors, real capital; also, a relatively large amount of content and too fast paced for 8th grade	XB0007
0	The content of the video explanation is not or only barely linked to prior knowledge	(No expression in the data)	

Notes on demarcation: What is not part of this criterion?

List of aspects that do not fit this criterion: Should be rated in:

The content is not connected to the everyday lives of the learners Relevance to the learners (L1)

Active engagement (L3).

Definition: Learners are encouraged to actively participate if the explainer video contains tasks that learners must complete during the video

Criteria: There is a call to pause or the video leaves a short time for the task
The number of tasks is adapted to the length of the video, and the video is not overloaded with tasks or does not contain any tasks. *

Task type: The task type is adapted to the topic and the learners

*Note: the appropriate number depends on the length of the video. Authors' estimate: 3 smaller tasks for 5 min is appropriate (for example: check prior knowledge, develop solution idea, check understanding)

Ratings:

A Rating Framework for the Quality of Video Explanations

Rating	Description	Examples for given reasons	Example Video IDs
3	Learners are fully encouraged to actively participate while watching the video explanation	(No reason necessary; in the example video, the use of tasks is ideal)	WU0103
2	Learners are mostly encouraged to actively participate while watching the video explanation	One break for the understanding task and one for the task with the juice example. A third break would have been useful at the beginning to activate previous knowledge	PK2215
1	Learners are partly encouraged to actively participate while watching the video explanation	Learners are asked at one point to pause the video to consider examples and check their understanding. This would also have been appropriate at other points	HA1819
0	Learners are not or only barely encouraged to actively participate while watching the video explanation	No request for active participation	XB0005

Notes on demarcation: What is not part of this criterion?

List of aspects that do not fit this criterion: Should be rated in:

At the end of the video a follow-up task is given Process

Direct addressing (L4).

Definition: Learners are addressed directly in the video explanation when the speaker uses second-person singular or first-person plural

Criteria: Quantity: How often are the relevant phrases used?

Ratings:

Rating	Description	Examples for given reasons	Example Video IDs
3	In the video explanation, the learners are always directly addressed	(No reason needed)	XM0003
2	In the video explanation, the learners are mostly directly addressed	The learner is addressed directly within the questions at various points and at the end of the video, e.g., "I hope you understood in this video..."	WU0103
1	In the video explanation, the learners are partly directly addressed	The learner is partly addressed directly, only at the beginning and end of the video, e.g., "Today we will talk about market forms..."	XB0002
0	In the video explanation, the learners are not or only barely directly addressed	The learner is not directly addressed in the entire video	RS1728

Notes on demarcation: What is not part of this criterion?

List of aspects that do not fit this criterion:	Should be rated in:
The learner's perspective is appropriately represented by an uninformed figure	Relevance to the learners (L1)

Use of examples (R1).

Definition:	Examples are used in the video explanation to promote learning if they clarify the domain content and represent the underlying principle
Criteria:	<p>Appropriate: The principle should be made clear by the examples/example; a typical case should be chosen (no exceptions, etc.)</p> <p>Number: The number is adapted to the topic; either one example is used throughout or different examples are used for different aspects</p> <p>Authentic and realistic: concrete companies, products, numbers, etc., instead of general phrasing</p>

Ratings:

Rating	Description	Examples for given reasons	Example Video IDs
3	In the video explanation, examples are always used to promote learning	(No reason needed)	XB0003
2	In the video explanation, examples are mostly used to promote learning	Examples are used well for the most part, but are sometimes rather general; or, no examples are used, only "suppliers" and "consumers" are referred to in the video (not specific enough)	HA1819
1	In the video explanation, examples are partly used to promote learning	No examples are used in the video except at the beginning	XM0005
0	In the video explanation, examples are not or only barely used to promote learning	No examples are used at all	XB0013

Notes on demarcation: What is not part of this criterion?

List of aspects that do not fit this criterion:	Should be rated in:
The example is not from the everyday lives of the learners	Relevance to the learners (L1)

Design of visualizations (R2).

Definition:	Visualizations are used in the video explanation to promote learning if they are technically correct and clearly designed
-------------	---

Criteria: Accuracy: no technical errors
 Design: no inaccuracies or incompleteness
 Clarity/number: The number and type of visuals are appropriate and the video is not cluttered or contains only a few visuals
 Step-by-step structure: For complex visualizations, clarity is also established through step-by-step construction and/or by highlighting information
 Contiguity (if it is text-image combination, the text and image should be arranged appropriately next to/on top of each other, not linked via a legend)
 Note: The criterion can be rated completely without an audio track—all aspects related to the fit of visualization should be rated in matching of visualizations and spoken text (R3)
 Note: If a video uses very few visualizations, this may affect the R2 and R3 ratings

Ratings:

Rating	Description	Examples for given reasons	Example Video IDs
3	In the video explanation, visualizations are always designed to promote learning	(No reason needed)	XB0004
2	In the video explanation, visualizations are mostly designed to promote learning	Mostly well designed; diagrams of company goals are not well designed as the axis labeling is missing	IY1506
1	In the video explanation, visualizations are partly designed to promote learning	Often very full and blurred; here, a step-by-step structure would have been very helpful (directly at the beginning but also at principles in the table, the table is unclear)	PK2215
0	In the video explanation, visualizations are not or only barely designed to promote learning	Entire video text-based; slides are also full of text, with no visualizations	XB0010

Notes on demarcation: What is **not** part of this criterion?

List of aspects that do not fit this criterion:	Should be rated in:
The visualizations do not focus on the most relevant information	Matching of visualizations and spoken text (R3)
The visualizations contain many unnecessary details	Matching of visualizations and spoken text (R3)
A comic strip or cartoon is used and not discussed	Matching of visualizations and spoken text (R3)

Matching of visualizations and spoken text (R3).

Definition: Visualizations and spoken content are used in the video explanation to promote learning when they are coordinated in terms of content and timing

Criteria:	Content fit: Visualization and spoken text fit together in terms of content (no general visualizations that are only remotely related to the topic) Temporal fit: Visualizations are in the picture/appear when the topic is being discussed Coherence: Visualizations are focused on the content that is essential for the topic (no unnecessary details in visualizations) Consistency: The same visualizations for the same content and different visualizations for different content Redundancy: The spoken text is usually included in the written text in keywords at most (definitions, legal texts, or tasks may be shown in full sentences) Note: If a video uses very few visualizations, this may affect the R2 and R3 ratings
-----------	---

Ratings:

Rating	Description	Examples for given reasons	Example Video IDs
3	In the video explanation, spoken text and visualizations are always linked in a way that promotes learning	(No reason needed)	WU0103
2	In the video explanation, spoken text and visualizations are mostly linked in a way that promotes learning	Redundancy: The spoken text is completely included in the image in one part of the video (especially in the checklist)	XB0016
1	In the video explanation, spoken text and visualizations are partly linked in a way that promotes learning	Visualization and spoken content are only partially linked, some visualizations are only linked to a limited extent or not at all with content (it is unclear, for example, why the comic is in the video, partly only speaking figures are visualized in the video, not the content they talk about)	XB0012
0	In the video explanation, spoken text and visualizations are not or only barely linked in a way that promotes learning	Spoken text is almost completely redundant with written information over the entire video if visualizations are used, they do not fit with the content	XB0010

Notes on demarcation: What is not part of this criterion?

List of aspects that do not fit this criterion:	Should be rated in:
In one visualization, an arrow is upside down	Design of visualizations (R2)
The slides/video are convoluted with visualizations	Design of visualizations (R2)

Comprehensive language (S1).

Definition:	The language of the video explanation is comprehensive if it is kept as simple as possible in terms of grammar and use of foreign words
Criteria:	No long subordinate clauses No non-technical foreign words or complex vocabulary Note: Necessary foreign words, such as oligopoly/monopoly/polypoly, for markets do not fall under this criterion. This is explicitly about linguistic elements that are not topic related

Ratings:

A Rating Framework for the Quality of Video Explanations

Rating	Description	Examples for given reasons	Example Video IDs
3	In the video explanation, the language is always comprehensive	(No reason needed)	WU0103
2	In the video explanation, the language is mostly comprehensive	Long sentences/sentence structure are not always flawless	XB0016
1	In the video explanation, the language is partly comprehensive	(No expression in the data)	
0	In the video explanation, the language is not or only barely comprehensive	(No expression in the data)	

Notes on demarcation: What is not part of this criterion?

List of aspects that do not fit this criterion:

Should be rated in:

The speaker mispronounced words

Precise language (S2)

Precise language (S2).

Definition: The language of the video explanation is precise when words are pronounced clearly and correctly, and speech tempo and pauses are used appropriately

Criteria:

- Accent- and dialect-free
- No verbal errors
- Fitting speech tempo
- Appropriate pauses
- Appropriate sound quality (easy to understand)

Ratings:

Rating	Description	Examples for given reasons	Example Video IDs
3	In the video explanation, the language is always precise	(No reason needed)	XM0001
2	In the video explanation, the language is mostly precise	One verbal error (is corrected immediately)	XB0005
1	In the video explanation, the language is partly precise	Some verbal errors, also inappropriate pauses	XB0011
0	In the video explanation, the language is not or only barely precise	Clear dialect, several verbal errors (also: laughing/coughing, etc.)	XB0014

Notes on demarcation: What is not part of this criterion?

List of aspects that do not fit this criterion:

Should be rated in:

The content of a sentence is not technically correct

Technical correctness (C1)

Structure (P1).

Definition:	The video explanation is well structured when the argumentation is coherent
Criteria:	Objective/topic/question at the beginning of the video Clear end—for example, by follow-up task, summary, or announcement of a new topic Questions that are raised at the beginning are also answered throughout the video Coherent structure of the argumentation, logical sequence of the explained content No jumps to other topics Consistent use of terminology (same terms for the same content)

Ratings:

Rating	Description	Examples for given reasons	Example Video IDs
3	The structure of the video explanation is always clear	(No reason needed)	PK2215
2	The structure of the video explanation is mostly clear	No clear end by follow-up task, summary, or announcement of a new topic; instead, a very abrupt end	XB0004
1	The structure of the video explanation is partly clear	The end is unclear, the introduction could be longer, some of the transitions are not ideal, partly unstructured	XB0014
0	The structure of the video explanation is not or only barely clear	(No expression in the data)	

Notes on demarcation: What is not part of this criterion?

List of aspects that do not fit this criterion:	Should be rated in:
The explanation is technically incorrect	Technical correctness (C1)
It is not clear why the topic/model is relevant for the domain	Technical completeness (C2)

Appendix 3: Spearman Correlation table of all criteria (significant correlations highlighted, boxes represent categories)

	C1	C2	L1	L2	L3	L4	R1	R2	R3	S1	S2	P1
C1	1.00 (NA)	-0.08 (n.s.)	0.16 (n.s.)	-0.08 (n.s.)	0.18 (n.s.)	-0.02 (n.s.)	0.23 (n.s.)	0.13 (n.s.)	0.10 (n.s.)	-0.06 (n.s.)	0.05 (n.s.)	0.26 (n.s.)
C2	-0.08 (n.s.)	1.00 (NA)	0.39 (p < 0,05)	0.24 (n.s.)	0.16 (n.s.)	-0.05 (n.s.)	0.39 (p < 0,05)	0.00 (n.s.)	0.10 (n.s.)	-0.03 (n.s.)	0.01 (n.s.)	0.25 (n.s.)
L1	0.16 (n.s.)	0.39 (p < 0,05)	1.00 (NA)	0.45 (p < 0,01)	0.19 (n.s.)	0.18 (n.s.)	0.81 (p < 0,001)	0.03 (n.s.)	0.23 (n.s.)	0.07 (n.s.)	0.09 (n.s.)	0.53 (p < 0,01)
L2	-0.08 (n.s.)	0.24 (n.s.)	0.45 (p < 0,01)	1.00 (NA)	0.19 (n.s.)	0.18 (n.s.)	0.46 (p < 0,01)	-0.08 (n.s.)	0.11 (n.s.)	-0.22 (n.s.)	0.13 (n.s.)	0.11 (n.s.)
L3	0.18 (n.s.)	0.16 (n.s.)	0.19 (n.s.)	0.19 (n.s.)	1.00 (NA)	0.51 (p < 0,01)	0.21 (n.s.)	0.09 (n.s.)	0.05 (n.s.)	-0.17 (n.s.)	-0.01 (n.s.)	0.05 (n.s.)
L4	-0.02 (n.s.)	-0.05 (n.s.)	0.18 (n.s.)	0.18 (n.s.)	0.51 (p < 0,01)	1.00 (NA)	0.21 (n.s.)	0.03 (n.s.)	0.19 (n.s.)	-0.12 (n.s.)	0.20 (n.s.)	0.16 (n.s.)
R1	0.23 (n.s.)	0.39 (p < 0,05)	0.81 (p < 0,001)	0.46 (p < 0,01)	0.21 (n.s.)	0.21 (n.s.)	1.00 (NA)	0.15 (n.s.)	0.42 (p < 0,05)	0.14 (n.s.)	0.11 (n.s.)	0.33 (p < 0,05)
R2	0.13 (n.s.)	0.00 (n.s.)	0.03 (n.s.)	-0.08 (n.s.)	0.09 (n.s.)	0.03 (n.s.)	0.15 (n.s.)	1.00 (NA)	0.69 (p < 0,001)	0.28 (n.s.)	0.20 (n.s.)	-0.01 (n.s.)
R3	0.10 (n.s.)	0.10 (n.s.)	0.23 (n.s.)	0.11 (n.s.)	0.05 (n.s.)	0.19 (n.s.)	0.42 (p < 0,05)	0.69 (p < 0,05)	1.00 (NA)	0.39 (p < 0,05)	0.34 (p < 0,05)	0.19 (n.s.)
S1	-0.06 (n.s.)	-0.03 (n.s.)	0.07 (n.s.)	-0.22 (n.s.)	-0.17 (n.s.)	-0.12 (n.s.)	0.14 (n.s.)	0.28 (n.s.)	0.39 (p < 0,05)	1.00 (NA)	0.42 (p < 0,05)	0.17 (n.s.)
S2	0.05 (n.s.)	0.01 (n.s.)	0.09 (n.s.)	0.13 (n.s.)	-0.01 (n.s.)	0.20 (n.s.)	0.11 (n.s.)	0.20 (n.s.)	0.34 (p < 0,05)	0.42 (p < 0,05)	1.00 (NA)	0.23 (n.s.)
P1	0.26 (n.s.)	0.25 (n.s.)	0.53 (p < 0,01)	0.11 (n.s.)	0.05 (n.s.)	0.16 (n.s.)	0.33 (p < 0,05)	-0.01 (n.s.)	0.19 (n.s.)	0.17 (n.s.)	0.23 (n.s.)	1.00 (NA)

Note This table shows the Spearman correlation coefficient for all criteria, exact *p* values can be taken from the additional material on the open science framework <https://doi.org/10.17605/OSF.IO/AKBWJ>

Appendix 4: Template—Rating sheet for video explanations

ID:

Topic:

Grade Level:

Criteria	Abbreviation	Rating (numeric, 0–3)	Reason
Technical correctness	C1		
Technical completeness	C2		
Relevance to the learners	L1		
Linking to prior knowledge	L2		
Active engagement	L3		
Direct addressing	L4		
Use of examples	R1		
Design of visualizations	R2		
Matching of visualizations and spoken text	R3		
Comprehensive language	S1		
Precise language	S2		
Structure	P1		

Additional aspects:

Appendix 5: Description of differences between theoretical framework and coding manual

One criterion that was theoretically derived, integration in the context, was not used in the coding manual because it had not been the student's task to describe how the video would be embedded in a lesson (and thus the criterion could not be rated). In contrast, active engagement was inductively added as a criterion in the category learner orientation. Although actively involving the learner is a frequent recommendation for video explanations, we had originally excluded this criterion from the theoretical framework due to our focus on content and the seemingly missing interactivity of video explanations. We then added the criterion again because in one of the five initially rated videos, the explainer asked the learners to pause the video for a task. Furthermore, in the process of developing the coding manual, the previously defined criterion relevance for the subject was discarded. Another originally dichotomous criterion (direct addressing of students) was expanded to a four-level criterion. For two criteria (technical completeness and linking to prior knowledge), we originally thought that raters should define their expectations in advance to rate the criteria according to how closely they conformed to their expectations. As the videos had multiple topics, however, clearly defining expectations in advance took too much time and was deemed impractical. We decided that raters should assess each criterion based on their expert knowledge and the economics curriculum with only the information about the grade level for which the video was created. Based on the expert feedback we received after the first version of the manual, we changed the classification of two criteria: direct addressing was classified as part of learner orientation (this criterion had been a subdimension of

language in the theoretical framework), and active engagement was also classified as part of learner orientation instead of process structure. Lastly, changes regarding the wording and further specifications were added to make the criteria more distinguishable.

Acknowledgements We would like to thank the preservice teacher students who produced the videos and agreed to their use for research purposes. We also thank Dr. Stefanie Findeisen and Patrick Schreyer for their feedback concerning the coding manual and Dr. Tim Kaiser, for his comments regarding an earlier version of this manuscript. Our gratitude further extends to the research assistant who supported us in the rating process.

Funding: Open Access funding enabled and organized by Projekt DEAL. The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

Data Availability The dataset, specifically the video ratings, as well as the codes used to produce the presented results are freely available in the Open Science Framework Repository, <https://doi.org/10.17605/OSF.IO/AKBWJ>

Declarations

Competing interests The authors have no relevant financial or non-financial interests to disclose.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Brame, C. J. (2016). Effective educational videos: Principles and guidelines for maximizing student learning from video content. *CBE Life Sciences Education*. <https://doi.org/10.1187/cbe.16-03-0125>
- Brown, G. A. (2006). Explaining. In O. Hargie (Ed.), *The handbook of communication skills*. Routledge.
- Charalambous, C. Y., Hill, H. C., & Ball, D. L. (2011). Prospective teachers' learning to provide instructional explanations: How does it look and what might it take? *Journal of Mathematics Teacher Education*, 14(6), 441–463. <https://doi.org/10.1007/s10857-011-9182-z>
- Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, 6(4), 284–290. <https://doi.org/10.1037/1040-3590.6.4.284>
- Delen, E., Liew, J., & Willson, V. (2014). Effects of interactivity and instructional scaffolding on learning: self-regulation in online video-based environments. *Computers & Education*, 78, 312–320. <https://doi.org/10.1016/j.compedu.2014.06.018>
- Findeisen, S. (2017). *Fachdidaktische Kompetenzen angehender Lehrpersonen: Eine Untersuchung zum Erklären im Rechnungswesen (Pedagogical content knowledge of preservice teachers: an empirical study of explanations in accounting) [Doctoral Dissertation, Universität Mannheim]*. Springer Fachmedien Wiesbaden.
- Fiorella, L., Stull, A. T., Kuhlmann, S., & Mayer, R. E. (2020). Fostering generative learning from video lessons: Benefits of instructor-generated drawings and learner-generated explanations. *Journal of Educational Psychology*, 112(5), 895–906. <https://doi.org/10.1037/edu0000408>
- Fiorella, L., van Gog, T., Hoogerheide, V., & Mayer, R. E. (2017). It's all a matter of perspective: Viewing first-person video modeling examples promotes learning of an assembly task. *Journal of Educational Psychology*, 109(5), 653–665. <https://doi.org/10.1037/edu0000161>
- Geelan, D. (2013). Teacher explanation of physics concepts: A video study. *Research in Science Education*, 43(5), 1751–1762. <https://doi.org/10.1007/s11165-012-9336-8>

- Hallgren, K. A. (2012). Computing inter-rater reliability for observational data: An overview and tutorial. *Tutorials in Quantitative Methods for Psychology*, 8(1), 23–34.
- Jaekel, A.-K., Scheiter, K., & Göllner, R. (2021). Distance teaching during the COVID-19 crisis: Social connectedness matters most for teaching quality and students' learning. *AERA Open*, 7, 233285842110520. <https://doi.org/10.1177/23328584211052050>
- Kay, R., & Rutenberg-Rozen, R. (2020). Exploring the Creation of Instructional Videos to Improve the Quality of Mathematical Explanations for Pre-Service Teachers. *International Journal of E-Learning and Distance Education*, VOL. 35, No. 1 (Special Issue on Technology and Teacher Education). <http://www.ijede.ca/index.php/ijede/article/view/1161/1795>
- Kay, R. (2014). Developing a framework for creating effective instructional video podcasts. *International Journal of Emerging Technologies in Learning*, 9(1), 22–30. <https://doi.org/10.3991/ijet.v9i1.3335>
- Köse, E., Taşlibeyaz, E., & Karaman, S. (2021). Classification of instructional videos. *Technology, Knowledge and Learning*, 26(4), 1079–1109. <https://doi.org/10.1007/s10758-021-09530-5>
- Kulgemeyer, C. (2018a). A framework of effective science explanation videos informed by criteria for instructional explanations. *Research in Science Education*. <https://doi.org/10.1007/s11165-018-9787-7>
- Kulgemeyer, C. (2018b). Towards a framework for effective instructional explanations in science teaching. *Studies in Science Education*, 54(2), 109–139. <https://doi.org/10.1080/03057267.2018.1598054>
- Kulgemeyer, C., & Peters, C. H. (2016). Exploring the explaining quality of physics online explanatory videos. *European Journal of Physics*, 37(6), 1–14. <https://doi.org/10.1088/0143-0807/37/6/065705>
- Kulgemeyer, C., & Schecker, H. (2013). Schülerinnen und Schüler erklären Physik-Modellierung, Diagnostik und Förderung von Kommunikationskompetenz im Physikunterricht [Students explain physics - modeling, diagnostics, and promoting communication skills in the physics classroom]. In M. Becker-Mrotzek, K. Schramm, E. Thürmann, & H. J. Vollmer (Eds.), *Ciando library: Vol. 3. Sprache im Fach: Sprachlichkeit und fachliches Lernen*. Waxmann.
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159–174. <https://doi.org/10.2307/2529310>
- Lee, H. S., & Anderson, J. R. (2013). Student learning: What has instruction got to do with it? *Annual Review of Psychology*, 64, 445–469. <https://doi.org/10.1146/annurev-psych-113011-143833>
- Leinhardt, G. (2001). Instructional explanations: A commonplace for teaching and location for contrast. In V. Richardson (Ed.), *Handbook of research on teaching*. American Educational Research Association.
- Low, R., & Sweller, J. (2014). The modality principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 227–246). Cambridge University Press.
- Marquardt, K. (2016). *Beurteilungsraster für Mathematik-Erklärvideos: Chancen, Grenzen und Durchführung einer Operationalisierung mittels Resultaten aus der Schulbuchforschung [A Framework for Video Explanations in Maths: Opportunities, Limitations, and Implementation of Operationalization Using Results from Textbook Research.]* [Magister Thesis, Wien University]. <http://othes.univie.ac.at/43166/1/44890.pdf>
- Mayer, R. E. (2014). Cognitive theory of multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning*. Cambridge University Press.
- Mayer, R. E. (2014). *The Cambridge handbook of multimedia learning*. Cambridge University Press. <https://doi.org/10.1017/CBO9781139547369>
- Mayer, R. E., & Fiorella, L. (2014). Principles for reducing extraneous processing in multimedia learning: Coherence, signaling, redundancy, spatial contiguity, and temporal contiguity principles. In R. E. Mayer (Ed.), *The Cambridge Handbook of multimedia learning* (pp. 279–315). Cambridge University Press.
- Mayer, R. E., Fiorella, L., & Stull, A. (2020). Five ways to increase the effectiveness of instructional video. *Educational Technology Research and Development*, 68(3), 837–852. <https://doi.org/10.1007/s11423-020-09749-6>
- Ring, M., & Brahm, T. (2020). Logical pictures in secondary economic education: Textbook analysis and teacher perception. *Research in Subject-Matter Teaching and Learning*, 3, 86–107.
- Schopf, C. (2020). Erklärvideos für den Unterricht gestalten - Empfehlungen aus Theorie, Empirie und Praxis [Developing Video Explanations for Instruction - Recommendations From Theory, Empirical Research and Practice]. In B. Greimel-Fuhrmann & R. Fortmüller (Eds.), *Wirtschaftsdidaktik: Den Bildungshorizont durch Berufs- und Allgemeinbildung erweitern. Festschrift für Josef Aff* (1st ed., pp. 61–78). Facultas.
- Schopf, C., Raso, A., & Kahr, M. (2019). How to give effective explanations: Guidelines for business education, discussion of their scope and their application to teaching operations research. *RISTAL-Research in Subject-Matter Teaching and Learning*, 2, 32–50.
- Schopf, C., & Zwischenbrugger, A. (2015). Verständliche Erklärungen im Wirtschaftsunterricht [Comprehensible explanations in business education]. *Zeitschrift Für Ökonomische Bildung*, 3, 1–31.

- Sevian, H., & Gonsalves, L. (2008). Analysing how scientists explain their research: A rubric for measuring the effectiveness of scientific explanations. *International Journal of Science Education*, 30(11), 1441–1467. <https://doi.org/10.1080/09500690802267579>
- Siegel, S. T., & Hensch, I. (2021). Qualitätskriterien für Lehrvideos aus interdisziplinärer Perspektive: Ein systematisches Review [quality criteria for educational videos from an interdisciplinary perspective - a systematic review]. In E. Matthes, S. T. Siegel, & T. Heiland (Eds.), *Lehrvideos – das Bildungsmedium der Zukunft? Erziehungswissenschaftliche und fachdidaktische Perspektiven [Instructional videos - the educational medium of the future? Educational science and subject didactic perspectives]* (pp. 254–266). Verlag Julius Klinkhardt.
- Sweller, J. (2020). Cognitive load theory and educational technology. *Educational Technology Research and Development*, 68(1), 1–16. <https://doi.org/10.1007/s11423-019-09701-3>
- van Gog, T. (2014). The signaling (or cueing) principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 263–278). Cambridge University Press.
- van Gog, T., Verveer, I., & Verveer, L. (2014). Learning from video modeling examples: Effects of seeing the human model's face. *Computers & Education*, 72, 323–327. <https://doi.org/10.1016/j.compedu.2013.12.004>
- Vazquez, J. J., & Chiang, E. P. (2014). A picture is worth a thousand words (at least): The effective use of visuals in the economics classroom. *International Review of Economics Education*, 17, 109–119. <https://doi.org/10.1016/j.iree.2014.08.006>
- Wittwer, J., & Renkl, A. (2008). Why instructional explanations often do not work: A framework for understanding the effectiveness of instructional explanations. *Educational Psychologist*, 43(1), 49–64. <https://doi.org/10.1080/00461520701756420>

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