Journal of Research in Science, Mathematics and Technology Education

https://jrsmte.com/

A Criteria-based Assessment Instrument to Assess the Quality of Explanatory Videos in Chemistry Education

Sarah Merk, Dominik Diermann, Katharina Forster & Jenna Koenen

Technical University of Munich, Germany

Abstract: The availability and use of instructional and explanatory videos are becoming increasingly important in educational contexts, e.g., due to advancing digitalization and the rising popularity of video platforms. Empirical findings from various subject areas and disciplines examine and postulate relevant quality criteria for these videos. To ensure the positive learning effects of the chosen videos, a qualitative assessment is required based on findings from media education, educational science, and (scientific) subject education. There is a significant research gap in the systematic quality assessment of explanatory videos in chemistry. As a result, an assessment instrument for recording and evaluating the quality of explanatory videos specifically for chemistry was developed. Based on a systematic literature search and a qualitative expert survey, 28 empirically founded quality criteria from media pedagogy, educational science, and science education were collected and validated based on N = 16 videos. The results of the expert assessments largely underline the quality and validity of the assessment instrument developed. The experts graded the given videos similarly and comparable to the quality assessment instrument, but, on average, they only used nine of the 28 identified criteria from the instrument for their video assessment, which speaks in favor of a more differentiated video evaluation by the assessment instrument. The instrument also includes the option of setting individual priorities and is widely applicable to different video formats (e.g., experimental videos) for the subject area of chemistry. Furthermore, all of the studies provide many indications of high user-friendliness, satisfaction, comprehensibility, and ease of use of the assessment instrument.

Keywords: *Explanatory videos; Chemistry lessons; Assessment instrument; Quality assessment.*
DOI: https://doi.org/10.31756/jrsmte.417SI

Introduction

Digitalization has triggered far-reaching changes in many areas of society in recent years. Education is no exception. Increasing digitalization offers many opportunities to expand and improve traditional teaching methods, particularly in scientific subjects such as chemistry (Jebe et al., 2019; Rummler & Wolf, 2012; Wolf, 2015). One established approach in this context is using explanatory and experimental videos. These short, often animated videos explain complex chemical concepts and processes in a clear and easy-to-understand way and are becoming increasingly popular in the education sector: users are enthusiastic about the simplicity and "low-threshold access to knowledge and educational content" (Honkomp-Wilkens et al., 2022, p. 495, translated) as well as the learning opportunities that become available through video platforms.

Studies show that platforms such as YouTube are no longer only used for entertainment purposes but have also become a kind of "visual encyclopedia" (Wolf, 2015, p. 121, translated) that covers both everyday and highly specialized knowledge. These videos are also becoming increasingly important as a "learning resource both for informal learning processes and for school or academic learning" (Wolf, 2015, p. 126, translated). Especially since the pandemic-related school closures, teachers are also increasingly turning to video-based learning materials in the classroom (Das Deutsche Schulportal, 2022). The use and self-production of explanatory videos are gaining importance, facilitated by technical progress and digitalization (Rummler & Wolf, 2012).

Given the increasing availability and use of explanatory videos in the education sector, it is crucial to assess their quality and distinguish between 'good' and 'bad' videos (Meller, 2017). However, this is often done intuitively and less according to formal and evaluated criteria. Scientifically validated catalogs of criteria and checklists could help to objectively assess the quality of explanatory videos (Findeisen et al., 2019; Meller, 2017). However, specific evaluation tools focusing on chemistry education are very rare or unsystematic (e.g., Harmer & Groß, 2021; 2022). This raises the question of how the quality of explanatory videos in chemistry can be assessed to make a well-founded selection of high-quality explanatory videos for teaching and learning processes. These questions will be examined in more detail below. The current state of research on quality criteria and assessment instruments for explanatory videos will be analyzed to develop a specific assessment instrument for the field of chemistry. This should ultimately help to further optimize chemical teaching and learning processes through high-quality digital content.

Theoretical Background

It is important to categorize and differentiate between various video formats used in teaching and learning. Depending on the educational orientation, complexity, and objective, different terms are often blurred or used synonymously in practice. Wolf's typology (2015) can create more clarity and distinguish between professionally and self-produced formats. Professional formats like educational films are didactically designed productions that specifically support learning processes. In contrast, documentary films depict events or facts but do not pursue a specific educational purpose. Self-produced formats are created by so-called "prosumers" (producing consumers) and comprise various types of video, including educational videos, instructional videos, explanatory videos, and performance videos.

Explanatory videos are usually created by non-professionals, such as students or teachers (Findeisen et al., 2019; Wolf, 2015). They are intended to create understanding and stimulate learning processes, but are less structured to support learning than instructional videos. Instructional videos, on the other hand, are more didactically prepared to achieve specific learning objectives. Performance videos, on the other hand, document simple actions or processes without any additional educational claim. This article, therefore, defines explanatory videos as the sum of these video types.

Explanatory videos are used by both learners and teachers in different contexts, independent of time and place. Learners receive content to prepare for exams or review and develop their subject knowledge. Some produce their videos as part of school or university projects and publish them on platforms like YouTube. Teachers use explanatory videos in lessons, for example, as an "eye-catcher" to arouse interest, substitute for lectures, or demonstrate new content. Overall, explanatory videos can take many forms depending on the objective and application. Evaluating their quality based on specific criteria is essential to ensure effective learning processes (Rüsseler et al., 2017).

It is already known and empirically proven from various perspectives that explanatory videos have a high motivational and learning potential, provided they are designed and produced to a high standard (cf. for large classes see Preston et al., 2010). On the other hand, however, they do not offer the opportunity for direct interaction, which can make comprehension more difficult (Balcke, 2022). Videos are naturally restricted to a limited section of reality, and their production can be time-consuming and problematic regarding copyright (Krammer & Reusser, 2005; Preston et al., 2010). These aspects should be considered when using and evaluating explanatory videos. Further advantages and educational reasons for using explanatory videos are summarized below.

Learning with Explanatory Videos as Information Processing

The influence of media on learning is often explained by models of information processing that emphasize the selection, organization, and integration of information (Dorgerloh & Wolf, 2020). According to Mayer's “Cognitive Theory of Multimedia Learning” (Mayer, 2014), working memory has separate visual and auditory information capacities. Simultaneously, using both channels (multimedia principle) and considering contiguity principles (spatial and temporal proximity of text and image) can make learning more effective. The “Cognitive Load Theory” (Sweller, 1988) supports this view by emphasizing the limited capacity of the working memory, whose extraneous cognitive load should be kept as low as possible through optimal media design. Here, well-designed explanatory videos can help: Thanks to their multimedia design, they can support information processing and stimulate learning processes (e.g., Preston et al., 2010; Tenberg, 2021).

Learning from a Model

The theory of observational learning shows that it is often easier to imitate observed actions (e.g., in a video) than to follow verbal instructions (Brass et al., 2000). Studies by Hommel and Stränger (1994) differentiate between process-like and result-oriented movement imitation. Both forms are important for learning with instructional videos, whereby success depends on how complex the action is, how good the video quality is, and what prior knowledge the learners have (Wolf, 2015).

Learning through Reflection and Analysis

More recent approaches focus on case-based, research-based reflection and analysis of teaching to promote a deeper understanding of teaching-learning processes (Krammer & Reusser, 2005). Here, aspects such as grasping the complexity of reality and promoting flexibility in thinking are essential. Repeated viewing and discussion of video content can support these learning processes. A deeper understanding can be achieved through critical analysis, reflection, or students' production of videos (e.g., Martin & Kelchner, 1998; Rummler & Wolf, 2012). However, Bloom's revised taxonomy of learning objectives (Bloom, 1984; Churches, 2008) shows that learning through reflection and analysis of videos goes far beyond simple imitation and requires additional educational measures to achieve higher learning objectives in terms of taxonomy levels (Rummler & Wolf, 2012).

Learning through Teaching

The concept of learning through teaching (Martin & Kelchner, 1998) emphasizes that creating own instructional videos deepens the learning process. The preparation, explanation, and review phases are crucial for video creation. In the preparation phase, students are assumed to learn more intensively to explain the content later (Renkl, 1997).

During the explanation phase, they have to reorganize their knowledge, which stimulates metacognitive processes, among other things. The third phase of potential queries can be supported by integrating explanatory videos into cooperative learning processes in the classroom. In addition, "learning through teaching" promotes communication in the technical language (Rummler & Wolf, 2012) and thus the corresponding communicative skills. However, specific media skills and competencies are required to criticize, create, or select videos for chemistry lessons to use them sensibly and in a way that promotes learning. Many meta-analyses and studies have endeavored to empirically develop quality criteria for scientific explanatory videos for several years. Media pedagogical (cf. Coşkun et al., 2021; Findeisen et al., 2019; Guo et al., 2014; Seethaler et al., 2020), educational science (cf. Kulgemeyer, 2018; Müller & Oeste-Reiß, 2019) and subject didactics (Beautemps & Bresges, 2021; Liu et al., 2021; Naggar, 2022; Sterzing et al., 2021) perspectives collected various criteria, which are described in more detail below. The presented research will separate these perspectives for better understanding, although a clear separation is usually quite difficult, as all approaches overlap and can provide important information on overall quality. This needs to be remembered in practical applications. The only exceptions are approaches dealing only with individual quality aspects, such as explanatory quality (cf. Kulgemeyer, 2018).

Collections of Criteria from a Media Educational Perspective

Findeisen et al. (2019) conducted a meta-analysis of 24 studies on the effect of design elements in explanatory videos. They identified five key design features: (1) interactive elements such as additional material or adjustment of the video speed; (2) the perspective of the camera, which should ideally be from the point of view of the person explaining; (3) the age of the person explaining, with older people often being attributed more expertise, (4) the video duration, which should ideally be less than six minutes, and (5) an aesthetic design that increases positive emotions and motivation. This last point is also stressed by, for example, Guo et al. (2014) or Heidig et al. (2015).

Guo et al. (2014) investigated how production decisions influence student engagement and provided recommendations on video duration, visibility of the explainer, and the emotional setting. Coşkun et al. (2021) provided design suggestions for OLPs (Online Learning Platforms) based on interviews and eye-tracking data, highlighting usability, feedback opportunities, and video design criteria. Seethaler et al. (2020) developed a checklist for the evaluation of science educational videos, including categories such as "content and sequencing" and "cognitive support".

Collections of Criteria from an Educational Science Perspective

Müller and Oeste-Reiß (2019) developed an evaluation tool for learning materials, including educational design, content, costs, media design, social aspects, and usability. They also introduced the "Storytelling" category for explanatory videos (including the way of speaking or the introductory question on the topic). Kulgemeyer (2018) focused on the quality of explanations in physics videos and identified the seven core ideas: "adaptation", "use of illustrative tools", “clarification of relevance”, “structure”, “precession and coherence”, “explanation of concepts”, and “embedding into the lesson”.

Collections of Criteria from a Scientific and Didactic Perspective

Beautemps and Bresges (2021) examined the quality of science teaching videos on YouTube and developed a checklist with 17 recommendations on video structure, presenter, reliability, and subject area. Marquardt (2016) created an assessment grid for mathematics explanatory videos that includes general, subject-didactic-content, subject-didactic-methodological, media-scientific-technical, and pedagogical criteria. Naggar (2022) collected quality criteria for mathematics explanatory videos by interviewing students, teachers, and video producers. Sterzing et al. (2021) developed design criteria for physics explanatory videos validated in an expert survey. Liu et al. (2021) created a list of criteria specifically for chemistry experiment videos, which includes the organization of the content, the execution of the experiment, and the video quality. However, this approach is very specific and provides little guidance on weighing the individual evaluation aspects. Generally, all subjects have differences regarding methods and educational goals, with chemistry emphasizing scientific inquiry and experimental approaches (e.g., Ahmad et al., 2023; de Jong et al., 2013; Jegstad, 2023; Nerdel, 2017; Sommer et al., 2018). As a result, there are different needs for educational videos that teach chemistry topics in the best way possible. Very few of the previously identified criteria (e.g., Marquardt, 2016; Naggar, 2022; Sterzing et al., 2021) are universal criteria that can be applied to all subjects; rather, some subject-specific criteria and considerations for the design of educational videos on chemistry.

Research Desideratum

This is where the research work presented in this article starts, addressing the research desideratum for a special catalog of criteria for assessing the quality of educational and explanatory videos for chemistry. Chemistry, with its subject-specific working methods and content (e.g., Ahmad et al., 2023; de Jong et al., 2013; Jegstad, 2023), brings some special features and opportunities, which should also be reflected in chemical explanatory videos. Therefore, the assessment instrument presented here aimed to create a comprehensive, multi-perspective collection of criteria specifically for teaching chemistry. This should enable a differentiated quality assessment of external and self-produced explanatory and experimental videos.

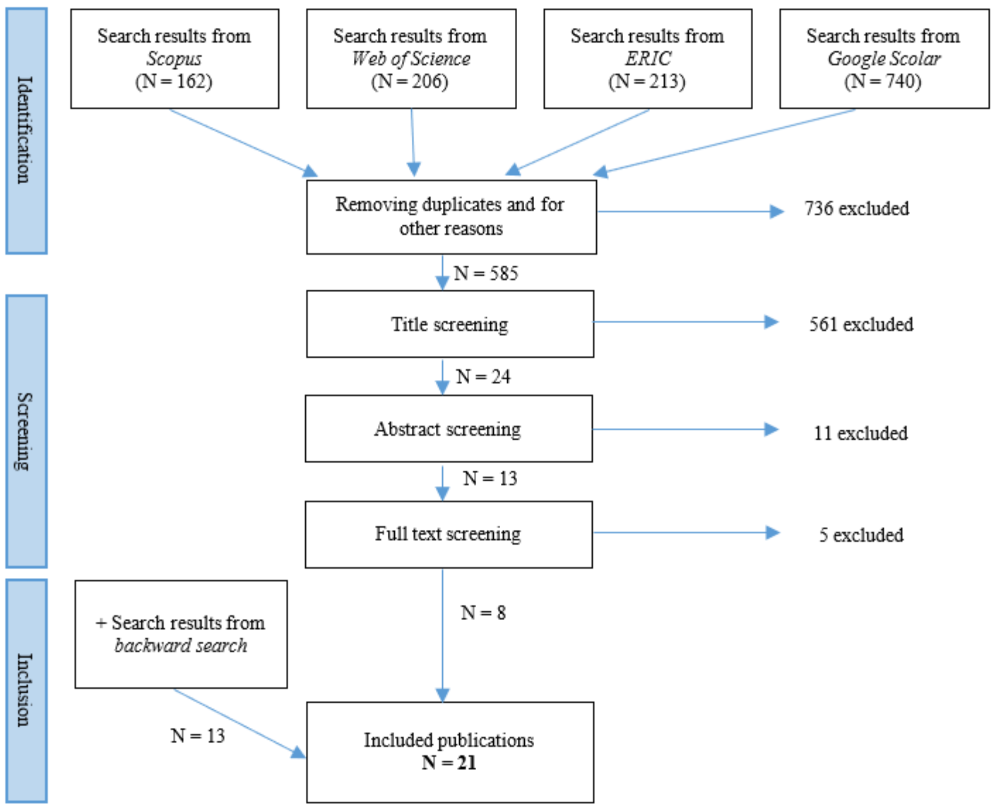
Thus, the presented research work aimed to answer the following research questions (RQ):
RQ 1: Which criteria are relevant for assessing the quality of educational and explanatory videos in chemistry?
RQ 2: What could a criteria-based assessment instrument look like that fulfills the content-related requirements of assessing chemistry education videos?
RQ 3: To what extent is the assessment instrument developed suitable for assessing the quality of explanatory videos in chemistry in a valid, applicable, and simple way?

Method

To develop and answer research questions 1 and 2, first, a systematic literature search (cf. Nordhausen & Hirt, 2020) with N = 585 hits (from Scopus, Web of Science, Google Scholar, and ERIC databases) was conducted, of which a final of $N = 21$ studies were included in the development of the 28 resulting assessment criteria in six categories. The entire search process is presented in the form of a PRISMA statement (cf. Moher et al., 2011; Page et al., 2021) as a "flow diagram" (see Figure 1).

Figure 1

PRISMA flow diagram (own adaptations based on Moher et al., 2011; Page et al., 2021)



Due to the screening, the final number of suitable publications was relatively low, so a backward search starting from the cited literature by the included works was additionally conducted to ensure that all relevant publications in the field were included.

The results of the included studies were then analyzed. The focus was on the quality criteria or design features of explanatory videos that were empirically collected and/or sufficiently validated in the respective studies. The identified quality criteria were first collected, including description and subcategories, checked for applicability, summarized, and divided into six areas (see Figure 2 and the following chapter). This resulted in the following six areas:

1. Science education and content-related quality criteria
2. Structural quality criteria
3. Science education and methodological quality criteria

4. Science education and pedagogical quality criteria
5. Media design and technical quality criteria
6. Personal-related quality criteria

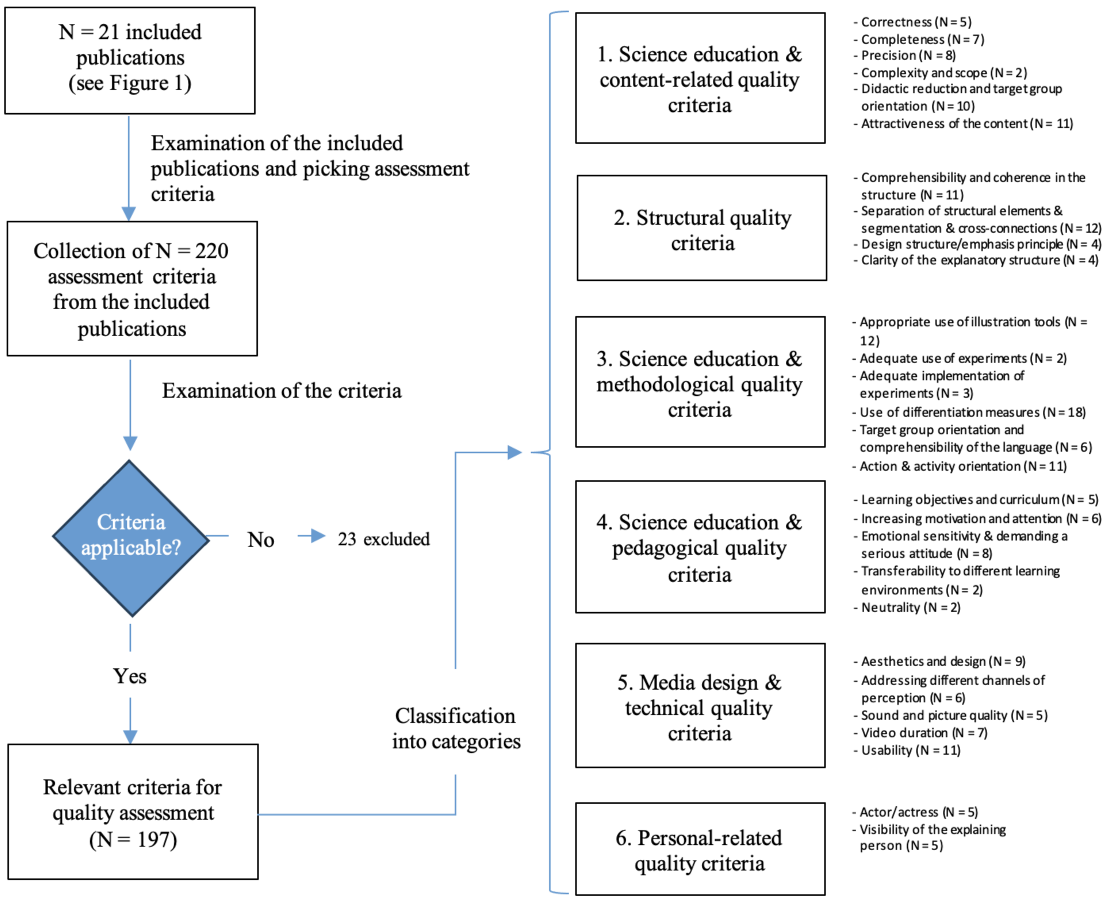
Areas 1, 3, 4, and 5 were taken from the assessment grid for math explanatory videos (see Marquardt, 2016). Areas 2 and 6 were added from other sources to add or specify structural and personal quality criteria. Figure 2 shows the identification, summary, and structure of the quality criteria for the later assessment instrument and their division into six criteria areas/categories.

To test whether the systematically created assessment instrument fulfills the self-imposed content-related and application-related requirements and to what extent it is understandable and user-friendly, a pilot validation test was carried out with $N = 2$ people with a chemistry teaching background and three videos; as a result of which the assessment instrument was again slightly adapted in terms of usability.

The instrument was then validated from a content and usage perspective in a subsequent expert survey with $N = 3$ chemistry education experts. As a first step, 16 freely available online explanatory videos of different types and quality levels were subjectively selected by an independent expert in chemistry education. These explanatory videos consider different didactic approaches and target groups. The experts watched the videos in the second step and independently assigned school grades from 1 to 6, with 1 representing excellent performance and 6 indicating poor performance. This step was carried out without using a standardized evaluation tool in order to allow an evaluation of the video quality based on individual assessments, which were rather unsystematic and without explicit support. In addition to assigning grades, the experts were asked to justify their evaluation criteria and rank the videos, contributing to the evaluation's transparency and comprehensibility. At the same time, the videos were rated, graded, and ranked by the first author of this article using the assessment instrument. The experts' ratings and results were then compared with the evaluation results using the assessment instrument. Thirdly, the experts rated the 28 criteria of the assessment instrument according to their importance on a scale of 'very important', 'important', to 'not important'. This step was used to check the relevance of the criteria. Finally, all results were summarised to assess the evaluation tool's comprehensibility, its application to different video formats, and the validity of the criteria.

Figure 2

Flow diagram for identifying, summarizing, and structuring the quality criteria with the number of empirical evidences for each assessment criterion on the right.



Results

Results and Assessment Categories of the Systematic Literature Search

First, the systematic literature search results on the quality assessment of explanatory videos are summarized (see research question 1). The analysis is divided into six main areas of quality criteria (see Figure 2). Empirical grounding of these criteria in supporting study results and meta-analyses can be found in the appendix (Table S1).

Science education and content-related quality criteria comprise 43 empirically determined aspects, which are divided into six sub-categories. “Correctness” (cf. Marquardt, 2016; Müller & Oeste-Reiß, 2019; Naggar, 2022) refers, in particular, to the technical appropriateness and consistency of the content, while “completeness” (cf. Marquardt, 2016; Müller & Oeste-Reiß, 2019; Naggar, 2022) assesses whether all the necessary technical information is available. “Precision” (cf. Coşkun et al., 2021; Kulgemeyer, 2018; Marquardt, 2016; Naggar, 2022; Sterzing et al., 2021) refers to the avoidance of superfluous details and the clear presentation of information. The “Complexity and scope” criteria require the content to be adapted to the target group. At the same time, “Didactic reduction” emphasizes adapting to the learners' prior knowledge and interests. Finally, the “Attractiveness of the content” (Harmer & Groß, 2021; 2022; Kulgemeyer, 2018; Müller & Oeste-Reiß, 2019; Naggar, 2022; Sterzing et al., 2021) also includes aspects such as the interdisciplinary treatment of topics and age-appropriateness.

Structural quality criteria comprise a total of 31 aspects, which are divided into four sub-categories relating to the clear and logical structure of the content (cf. Marquardt, 2016; Müller & Oeste-Reiß, 2019; Seethaler et al., 2020; Sterzing et al., 2021). This includes “Comprehensibility and coherence in the structure” (Beautemps & Bresges, 2021; Kulgemeyer, 2018; Marquardt, 2016; Müller & Oeste-Reiß, 2019; Naggar, 2022), “Separation of structural elements & segmentation & cross-connections” (Marquardt, 2016; Sterzing et al., 2021), “Desing structure/emphasis principle” (Marquardt, 2016; Sterzing et al., 2021), and “Clarity of the explanatory structure” (Kulgemeyer, 2018; Müller & Oeste-Reiß, 2019; Sterzing et al., 2021). The structure should include an introduction, clear topic naming, and a structured outline, while separating structural elements such as definitions and examples is important. Visual and structural principles should support the design structure and the clarity of the explanations.

Science education and methodological quality criteria take 52 aspects into account, which can be divided into six sub-categories. These relate to the use of “Appropriate illustration tools” like models (cf. Coşkun et al., 2021; Kulgemeyer, 2018; Liu et al., 2021; Müller & Oeste-Reiß, 2019; Naggar, 2022; Seethaler et al., 2020; Sterzing et al., 2021) and the “Adequate use (Liu et al., 2021) and implementation of experiments” (Kulgemeyer, 2018; Liu et al., 2021; Sterzing et al., 2021). Important further aspects are “Differentiation measures” (e.g., Kulgemeyer, 2018; Marquardt, 2016; Naggar, 2022) and the individual “Target group orientation”. “Comprehensible language” (cf. Liu et al., 2021; Marquardt, 2016; Naggar, 2022; Seethaler et al., 2020) and an “Action-oriented design” promote active learning by the audience (Fiorella et al., 2017; Kulgemeyer, 2018; Liu et al., 2021; Müller & Oeste-Reiß, 2019; Naggar, 2022; Sterzing et al., 2021).

Science education and pedagogical criteria comprise 23 aspects, divided into five sub-categories, which relate to the communication of “Learning objectives and curriculum” (Marquardt, 2016; Müller & Oeste-Reiß, 2019), the “Increasing motivation and attention” (Müller & Oeste-Reiß, 2019), “Emotional sensitivity” (Marquardt, 2016; Naggar, 2022) of learners, “Transferability” (Müller & Oeste-Reiß, 2019) and the “Neutrality” (Marquardt, 2016; Müller & Oeste-Reiß, 2019) of the content. The videos should formulate clear learning objectives, promote interest and attention, and be usable in different learning contexts. In the relevant literature sources, the possibilities mentioned include, for example, relevance to the real world (Naggar, 2022), practical relevance (Müller & Oeste-Reiß, 2019), or clarification of the relevance for the target group (Seethaler et al., 2020; Sterzing et al., 2021). A balance between entertainment and serious discussion also seems advisable (cf. Beautemps & Bresges, 2021; Guo et al., 2014; Heidig et al., 2015; Marquardt, 2016; Müller & Oeste-Reiß, 2019; Naggar, 2022; Plass et al., 2014).

Media design and technological criteria include 38 aspects, divided into five further sub-categories, which are “Aesthetics and design” (e.g., Müller & Oeste-Reiß, 2019; Naggar, 2022; Sterzing et al., 2021), “Addressing different channels of perception” (Marquardt, 2016; Naggar, 2022; Sterzing et al., 2021), “Sound and picture quality” (Coşkun et al., 2021; Marquardt, 2016; Naggar, 2022), “Video duration” (Beautemps & Bresges, 2021; Coşkun et al., 2021; Müller & Oeste-Reiß, 2019; Naggar, 2022), and “Usability” (Müller & Oeste-Reiß, 2019; Naggar, 2022). The design should be aesthetic and straightforward (cf. Coşkun et al., 2021; Müller & Oeste-Reiß, 2019; Sterzing et al., 2021), appeal to different channels of perception (cf. Marquardt, 2016; Naggar, 2022; Sterzing et al., 2021), and have good sound and picture quality. The video duration should be as short as possible (see discussion), and its usability should be user-friendly.

Person-related criteria comprise ten aspects, which are divided into two sub-categories: the “Actor/actress” or the explainer (Beautemps & Bresges, 2021; Coşkun et al., 2021; Müller & Oeste-Reiß, 2019; Naggar, 2022) and the “Visibility of the explaining person” (e.g., Chen & Wu, 2015; Guo et al., 2014). The presence of the explainer should support learning and not distract from the content. Chen and Wu (2015) show that visible explainers motivate higher learning engagement and that certain production styles, such as "lecture capture", are more suitable than others.

Presentation of the Assessment Instrument

In summary, this results in 28 identified criteria, representing the criteria of the developed assessment instrument (see Figure 3). This formed a comprehensive basis for evaluating the quality of explanatory videos, which can be assessed methodically and in terms of content in the various areas, and thus provides an answer to research question 1. To ensure the individual use of the assessment instrument, an additional blank line was added (criterion 29), in which the user can add their own criterion and include it in the assessment.

The criteria were formulated as statements to accurately describe the requirements. A more detailed explanation and description of each criterion, including a localization in the literature and underlying research works, can be found in the appendix (Table S1).

To achieve the self-imposed content and application-related requirements (research question 2), among other things, a rating scale (0 to 4) is used to characterize a video and rate the extent of the individual categories. For the assessment instrument, the qualitative characteristics of each quality criterion can be precisely assessed. This allows the overall assessment to be transparent, comprehensible, and objective. A weighting factor (0 to 2) was assigned to each criterion to consider the individual significance of individual aspects regarding the overall assessment (cf. Korossy, 2022). Therefore, each assessment instrument user can individually adjust the instrument to the type of video (for example, explanatory or experimental video). The resulting individual rating of the criteria represents an added value of the tool, making it versatile and adaptable to different requirements and contexts.

Figure 3

Assessment instrument to evaluate the quality of explanatory videos in chemistry

A criteria-based assessment instrument to assess the quality of explanatory videos in chemistry education

Quality criteria for explanatory videos in chemistry			Weighting factor			Criterion score/extent					Score	Max. score
			0	*1*	*2*	0	1	2	3	4		
Science education and content-related quality criteria	1	Correctness									0	0
	2	Completeness									0	0
	3	Precision									0	0
	4	Complexity and scope									0	0
	5	Didactic reduction and target group orientation									0	0
	6	Attractiveness of the content									0	0
Structural quality criteria	7	Comprehensibility and coherence in the structure									0	0
	8	Separation of structural elements & segmentation & cross-connections									0	0
	9	Design structure/emphasis principle									0	0
	10	Clarity of the explanatory structure									0	0
Science education and methodological quality criteria	11	Appropriate use of illustration tools									0	0
	12	Adequate use of experiments									0	0
	13	Adequate implementation of experiments									0	0
	14	Use of differentiation measures									0	0
	15	Target group orientation and comprehensibility of the language									0	0
	16	Action and activity orientation									0	0
Science education and pedagogical quality criteria	17	Learning objectives and curriculum									0	0
	18	Increasing motivation and attention									0	0
	19	Emotional sensitivity & demanding a serious attitude									0	0
	20	Transferability to different learning environments									0	0
	21	Neutrality									0	0
Media design and technical quality criteria	22	Aesthetics and design									0	0
	23	Addressing different channels of perception									0	0
	24	Sound and picture quality									0	0
	25	Video duration									0	0
	26	Useability									0	0
Personal-related quality criteria	27	Actor/actress									0	0
	28	Visibility of the explaining person									0	0
Own criterion	29	...									0	0

Weighting factor	
0	Not important
1	Important
2	Very important
Criterion score/extent	
0	Not at all
1	Rather not
2	More or less
3	Rather
4	Completely

Score	
Max. score	
Overall score (in %)	

* calculated automatically by putting "x" in the instrument cells

Grade	Overall score in %
1 (very good)	100-85
2 (good)	85-70
3 (acceptable)	70-55
4 (sufficient)	55-40
5 (flawed)	40-25
6 (bad)	25-0

For the overall evaluation of an explanatory video, the score achieved for each criterion is calculated first. To do this, the characteristic value (0 to 4) is multiplied by the respective weighting factor:

Score achieved per criterion = weighting factor x characteristic value

The maximum achievable score for a criterion is then calculated by multiplying the weighting factor by the best possible characteristic value (4).

Maximum achievable score per criterion = weighting factor x 4

The sum of the achievable scores is then compared with the maximum achievable score and expressed as a percentage.

Total scores achieved per criterion / total maximum achievable scores per criterion x 100% = total score in %

The final assessment instrument is available as an Excel spreadsheet in the digital appendix for the automated calculation of the overall score and a final grade (evaluation key in Figure 3).

Results of the Expert Survey

The expert survey for the assessment instrument's evaluation focuses on the instrument's validity, comprehensibility, and applicability. The results show both similarities and differences in the assessments of experts and the assessment instrument, providing insights into the assessment criteria and their relevance as well as the comprehensibility and applicability of the assessment.

Comparison of Video Ranking

The comparison of the experts' ratings and the assessment instrument provides a detailed overview of the individual assessments of the evaluation criteria and the overall ranking of the videos. Table 1 illustrates the video rankings, starting with the top-rated videos (highest ranks, meaning best assessment result and lowest number). The videos were color-coded for clarity. This illustration shows similarities, as well as some differences in the ratings.

Table 1

Comparison of the video rankings: each video is shown in an individual color

Rank	Expert ranking 1	Expert ranking 2	Expert ranking 3	Assessment instrument ranking
1	Video 10	Video 11	Video 1	Video 10
2	Video 4	Video 2	Video 7	Video 11
3	Video 2	Video 10	Video 11	Video 1
4	Video 7	Video 4	Video 4	Video 4
5	Video 16	Video 1	Video 2	Video 14
6	Video 14	Video 7	Video 10	Video 7
7	Video 11	Video 9	Video 13	Video 2
8	Video 1	Video 12	Video 9	Video 5
9	Video 15	Video 5	Video 15	Video 9
10	Video 12	Video 3	Video 3	Video 12
11	Video 9	Video 8	Video 12	Video 8
12	Video 13	Video 14	Video 14	Video 3
13	Video 3	Video 15	Video 16	Video 13
14	Video 5	Video 13	Video 5	Video 15
15	Video 8	Video 16	Video 8	Video 16
16	Video 6	Video 6	Video 6	Video 6

One significant result is the consistent rating of video V6 as the worst video by all raters and the instrument. Further similarities can be seen for videos V4, V10, and V2, which were consistently rated better than other videos, such as V8, V3, and V13. Video V4 stands out as it was rated as the fourth-best video by three of the four raters, which shows a certain consistency in the ratings. Table 2 indicates that the ratings vary from grade 1 to grade 6, with the assessment rating tending to result in better grades than the expert ratings. Overall, the similarities between chemistry education experts' rankings and the results from the developed assessment instrument (see Tables 1 and 2) prove the good content validity of the instrument, as very experienced and competent educational researchers achieve similar results to the assessment instrument.

Table 2

Comparison of the video rankings

Rank	Expert ranking 1	Expert ranking 2	Expert ranking 3	Assessment instrument ranking
1	Video 10	Video 11	Video 1	Video 10
2	Video 4	Video 2	Video 7	Video 11
3	Video 2	Video 10	Video 11	Video 1
4	Video 7	Video 4	Video 4	Video 4
5	Video 16	Video 1	Video 2	Video 14
6	Video 14	Video 7	Video 10	Video 7
7	Video 11	Video 9	Video 13	Video 2
8	Video 1	Video 12	Video 9	Video 5
9	Video 15	Video 5	Video 15	Video 9
10	Video 12	Video 3	Video 3	Video 12
11	Video 9	Video 8	Video 12	Video 8
12	Video 13	Video 14	Video 14	Video 3
13	Video 3	Video 15	Video 16	Video 13
14	Video 5	Video 13	Video 5	Video 15
15	Video 8	Video 16	Video 8	Video 16
16	Video 6	Video 6	Video 6	Video 6
Ø	2,9	2,5	3,3	2,1

Grade 1 (best): light green - Grade 2: dark green - Grade 3: yellow –
Grade 4: orange - Grade 5: light red - Grade 6 (worst): dark red

When evaluating the *N = 16* videos, it becomes quantitatively clear that the assessment evaluation (Ø 2.1) tends to result in better grades than the three individual expert evaluations (Ø 2.9/2.5/3.3). Accordingly, the overall average of the assessment evaluation is 0.77 grade points below the average of the expert evaluations. These differences indicate that the assessment evaluation is slightly more generous in its grading overall than the expert evaluations.

Criteria Comparison

The assignment of the experts' free-text explanations for comparison with the assessment criteria of the assessment instrument was based on their respective descriptions (see appendix, Table S1). Specific formulations by the experts were clustered into superordinate terms and assigned to the 28 criteria identified in the literature research (if possible).

A total of 23 out of 28 quality criteria were mentioned in the experts' assessments. A complete overview of the criteria mentioned, their assignment to the assessment instrument, and the importance attributed to the criterion by all three experts can be found in the appendix (Table S2). The most frequently mentioned criteria were "Sound and picture quality" (58 mentions), "Adequate implementation of experiments" (57 mentions), "Comprehensibility and coherence in the structure" (46 mentions), "Aesthetics and design" (40 mentions) and "Correctness" (26 mentions). Criteria such as "Completeness" (14 mentions) and "Attractiveness of the content" (8 mentions) were mentioned less frequently (see Table 3). Some criteria, such as "Learning objectives and curriculum", " Emotional sensitivity & demanding a serious attitude", and "Transferability to different learning environments" were not mentioned in the experts' assessments, which may indicate that these criteria are less relevant.

Table 3

Ranking of the criteria according to the frequency with which the experts in the video assessment mentioned them

Rank	Number of mentions	Assigned assessment criterion	Expert assessment (averaged)
1	58	Sound and picture quality	Very important
2	57	Adequate implementation of experiments	Very important
3	46	Comprehensibility and coherence in the structure	Very important
4	40	Aesthetics and design	Very important
5	26	Correctness	Very important
6	22	Increasing motivation and attention	Important
7	20	Appropriate use of illustration tools	Very important
8	17	Addressing different channels of perception	Very important
9	16	Actor/actress	Not important
10	16	Target group orientation and comprehensibility of the language	Very important
11	15	Clarity of the explanatory structure	Very important
12	14	Completeness	Important
13	10	Separation of structural elements & segmentation & cross-connections	Very important
14	8	Attractiveness of the content	Very important
15	7	Action & activity orientation	Important
16	7	Didactic reduction and target group orientation	Very important
17	5	Design structure/emphasis principle	Very important
18	5	Adequate use of experiments	Very important
19	5	Complexity and scope	Very important
20	4	Video duration	Important
21	4	Precision	Important
22	3	Visibility of the explaining person	Not important
23	1	Use of differentiation measures	Not important
-	0	Learning objectives and curriculum	Very important
-	0	Emotional sensitivity & demanding a serious attitude	Important
-	0	Transferability to different learning environments	Important
-	0	Neutrality	Important
-	0	Usability	Very important

Relevance Assessment of Criteria by the Experts

The additional relevance assessment of the previously unknown 28 criteria of the instrument by the three experts shows that only the criteria "Actor/actress", "Visibility of the explaining person", and "Use of differentiation measures" were rated as "not important". In contrast, most criteria were described as "very important" (see Table 3 and appendix Table S2). In addition, a discrepancy can be seen between the attributed importance and the actual consideration in the respondents' written evaluations. For example, all experts rated "Usability " as "very important" but it was never mentioned in the assessments. In contrast, criteria such as "Completeness", which the experts only rated as "important" (middle of the scale), were mentioned more frequently in the evaluation of the videos. This indicates that specific criteria may not have been sufficiently considered in their evaluation or that the experts were not sufficiently aware of the criteria - or could not have been aware of them due to their number, complexity, and variety.

Comprehensibility of the Evaluation with the Instrument

The comprehensibility of the evaluation using the assessment instrument compared to the expert evaluations was examined quantitatively and qualitatively (see research question 2). The quantitative analysis shows that the three experts use an average of between three and 15 assessment aspects per video, with the average number of considered aspects being nine. This contrasts with the evaluation using the assessment instrument, which consistently covers all 28 criteria. This difference indicates that the assessment instrument enables a more detailed and consistent evaluation. Qualitatively, the analysis of the expert evaluations shows that often, no clear focus or reasons for the grading were given. This indicates that the assessment instrument evaluations are possibly more comprehensible, as they are based on a clearly defined and extensive catalog of criteria. The consistent application of the 28 criteria in the quality assessment and their explanatory descriptions contribute to the transparency and comprehensibility of the evaluation.

Applicability of the Assessment Instrument

The investigation of the applicability of the assessment instrument for different video formats and levels of professionalism (see video selection) shows that the instrument can be used consistently for different video types. On average, videos with experiments are rated around 1.1 (expert assessment) or 1.2 grade points (assessment instrument) better than videos without experiments. Furthermore, the results show no significant differences in quality between YouTube videos and videos from more reputable educational sources (e.g., university homepages). The ratings are close to each other (differences of 0.4 grade points for expert ratings and 0.2 grade points for assessment ratings). These similarities confirm the instrument's validity and underline the legitimacy of the instrument for evaluating different video formats (in particular, experimental videos) from different subject areas and levels of professionalism.

Discussion

In summary, a new assessment instrument was developed, largely legitimized by the systematic literature search (or the empirical data obtained) and expert evaluations. It enables a valid, detailed, and consistent evaluation in comparison, taking various aspects from different disciplines into account. Besides, it allows students and educators to put an individual focus. The results show that the instrument is suitable for different video formats and videos of varying levels of professionalism. Moreover, the consistent and comprehensible results provide a reliable basis for the quality assessment of explanatory videos. Overall, this research confirms the assessment instrument's validity and applicability as an effective tool for evaluating explanatory videos in chemistry and chemistry education.

Systematic Literature Search

The strictly systematic literature search made it possible to identify relevant empirical research contributions on the quality of explanatory videos, from which well-founded assessment criteria could then be obtained. Despite specific measures, it is possible that not all relevant literature could be identified (e.g., grey literature). Finally, it can be stated that $N = 197$ quality criteria relevant to the assessment of educational video quality were summarized to $N = 28$ quality criteria. Various findings related to teaching and learning processes, as well as, for example, safety aspects during experimentation (Liu et al., 2021) or the attractiveness of experiments, were considered to ensure the broadest possible assessment. The analysis showed that some criteria established in other subject areas appear transferable to chemistry videos, but have not been sufficiently scientifically tested, which could therefore be critically discussed. In some cases, different sources made contradictory claims about the quality of specific criteria, which also had to be considered (e.g., "Video duration"). For this reason, formulations/explanations were chosen for the categories (see Table S1) that allow the assessing person to take their personal feelings and priorities into account (e.g., "The video duration is appropriate (recommendation 3-15 min)"). Furthermore, the evaluation of the "specification of sources" (Beautemps & Bresges, 2021) showed that although many consider this criterion necessary, a significant minority rated it as less relevant. In addition, individual quality criteria such as "Transferability to different learning environments" and "Target group orientation and comprehensibility of the language" can be critically discussed. While Müller and Oeste-Reiß (2019) call for the flexibility of the didactic design for different learning environments, other studies (Kulgemeyer, 2018; Naggar, 2022) emphasize the need to adapt content to specific target groups. Both approaches were integrated into the evaluation criteria for different usage contexts. Generally, some criteria might be impossible to rate since there might be no information in some cases (e.g., “Usability”), or they might not appear in a video (e.g., “Adequate implementation of experiments”). Here, the weighting factor could simply be set to zero. Naturally, some criteria are better and more investigated than others, which is not reflected in the final assessment instrument.

Quality Assessment of the Assessment Instrument

The literature review resulted in a consolidated list of 28 quality criteria divided into six areas. The criteria cover various quality aspects, including subject educational, structural, methodological, pedagogical, media science, technical, and personal dimensions (cf. Mayer & Chandler, 2001; Kulgemeyer, 2018), making it unique in comparison to previous approaches (e.g., Seethaler et al., 2010). Moreover, the instrument particularly takes the field of chemistry education into account, which was very rarely done before. Nearly all criteria are comprehensively empirically proven, based on established and proven collections of quality criteria for explanatory videos from different subject areas and formats. Combining several previous studies made it possible to close gaps (e.g., due to missing criteria), meaning that the assessment instrument can be considered complete and comprehensive overall. Another reason why the collected quality criteria are justified is shown by the comparison with further and more general educational findings and theories on learning with multimedia, such as Mayer's design principles (cf. Mayer, 2014).

However, it should be noted that the compiled list also contains criteria that have been examined, justified, and differentiated to varying degrees in the literature. Nevertheless, these should be adequately supported by the existing quantity and quality of empirical evidence on the assessment criteria for explanatory videos in chemistry. The final quality criteria selection aimed to find a compromise between a possible lack of differentiation and practical applicability. The results of the study indicated that this compromise has been successful. Still, individual users might find the criteria list too little or too much differentiated and could criticize the number of categories.

In evaluating the $N = 16$ videos, the results in terms of the order from the best to the worst video rating showed many similarities between experts and the instrument's results (see Tables 1 and 2).

Overall, the grades achieved with the assessment instrument are better than the grades given by the experts, which may be due to the different focus of the assessment. While the experts seem to rely on individual and maybe more unsystematic applied or subjective criteria, the developed assessment instrument offers the possibility to evaluate videos in a differentiated, systematic way and based on a wider range of criteria. This could also mean that non-guided assessments focus on the negative aspects and forget to consider the positive ones.

The research shows that the instrument is suitable for both videos with and without experiments, as the differences in grades between the two formats are minor. Compared to YouTube videos, a greater difference in the expert rating was found when assessing videos from well-known educational sources. This indicates a possible unconscious influence during the expert validation study (as the experts did not know the video sources).

Expert Validation

The different video formats also represent a point of criticism of the methodology, as the experts also noted difficulties in evaluating and ranking the videos. This was because the videos were only comparable to a limited extent due to different video formats (see Wolf, 2015), different subject areas, and the associated learning objectives. A more systematic selection of videos with the same topics and formats would have been necessary to test the instrument's validity more precisely. An even higher number of assessed videos would have been beneficial as well. Additionally, the quite low number of experts (only from Germany) poses a limitation for the presented study and its results in terms of generalizability.

The expert survey results show that the tool enables a very differentiated assessment of various aspects of quality: on average, experts use around nine criteria to justify their video assessment, three times fewer than the number of criteria contained in the assessment instrument. This could indicate that the assessment instrument includes more categories than are required for a comprehensive quality assessment. However, the systematic approach to the development of the instrument and the empirical confirmation of all the criteria included argue against this. It is, therefore, more likely that the experts did not consider all relevant criteria in their non-directed assessment. One further methodological point of discussion in the expert evaluation was the deliberate decision to compare videos of different formats and topics. This can be seen as a legitimate criticism, as comparing videos with similar themes and formats would potentially provide more accurate results. This aligns with the recommendations of Liu et al. (2021) on the need for a systematic assessment of instructional videos to obtain valid results. However, the assessment instrument was developed to be universally applicable, and it is also simulated by the assessment of videos created by learners. This practice reflects the actual teaching context, as suggested by Riedl (2008), who emphasizes the relevance of realistic assessment scenarios for educational research.

Overall, however, the piloting and expert survey results confirmed the empirical basis of the selected criteria (research question 1) and their suitability for the developed assessment instrument (e.g., through the additional relevance assessments or comparable evaluation patterns/results). The instrument fulfills both content-related and application-related requirements (research question 2) and was identified as a valid, universally applicable method for a wide-ranging assessment of the quality of chemistry videos (research question 3).

Differences in the Ratings of Chemistry Explanatory Videos

The minor deviations between the assessments of experts and the assessment instrument could be due to individual assessment styles, personal priorities, or different diagnostic skills, as Wolf (2015) noted. The average score difference between various video formats (videos with and without experiments) is relatively small, according to both the experts and the assessment instrument, which shows a certain versatility of the tool. This responds to the statements by Wolf (2015), who emphasizes the difficulty of comparing different video types. To reduce these influencing factors, concrete role attributions for the experts could be helpful, for example, by simulating the assessment of videos in a teaching context. In summary, the results confirm that the assessment instrument can be applied to different subject areas, video formats, and levels of professionalism.

Possible Uses of the Assessment Instrument

To promote effective teaching and learning processes, an assessment instrument must be possible and easy to use, and suitable explanatory videos must be selected, whether for individual learning, usage in class, or as a basis for reflection on self-produced video formats. The presented novel tool comprehensively differentiates and records the quality of explanatory videos on chemistry, enabling a comparable evaluation of these videos. It can contribute to promoting various media skills among teachers and students: it addresses multimedia design by producing high-quality videos and the critical selection and reflection of chemical explanatory videos. In addition, the tool can strengthen the diagnostic skills of active or future teachers by enabling them to deal with relevant quality criteria during their studies and promoting topic-related awareness. Teachers can also benefit from this tool when assessing performance (e.g., if students create videos). However, the assessment instrument does not replace individual quality considerations, which should be made independently by users, creators, or evaluators. This applies in particular to the individual setting of priorities. Therefore, the previously presented possibility of personal weighting of the assessment criteria/categories can be seen as a particular and innovative advantage compared to previous approaches, as users can adapt the tool to the situation and target group. Based on the expert assessments of the relevance of the 28 categories (see Tables 3 and S2), a suggested weighting for explanatory videos on chemistry could be created (see Figure 4).

Figure 4

Proposed weighting of the assessment criteria of the assessment instrument based on the expert assessments

Quality criteria for explanatory videos in chemistry			*Weighting factor*		
			0	*1*	*2*
Science education and content-related quality criteria	1	Correctness			X
	2	Completeness			X
	3	Precision		X	
	4	Complexity and scope		X	
	5	Didactic reduction and target group orientation		X	
	6	Attractiveness of the content		X	
Structural quality criteria	7	Comprehensibility and coherence in the structure			X
	8	Separation of structural elements & segmentation & cross-connections			X
	9	Design structure/emphasis principle		X	
	10	Clarity of the explanatory structure			X
Science education and methodological quality criteria	11	Appropriate use of illustration tools			X
	12	Adequate use of experiments		X	
	13	Adequate implementation of experiments			X
	14	Use of differentiation measures		X	
	15	Target group orientation and comprehensibility of the language			X
	16	Action and activity orientation		X	
Science education and pedagogical quality criteria	17	Learning objectives and curriculum	X		
	18	Increasing motivation and attention			X
	19	Emotional sensitivity & demanding a serious attitude	X		
	20	Transferability to different learning environments	X		
	21	Neutrality	X		
Media design and technical quality criteria	22	Aesthetics and design			X
	23	Addressing different channels of perception			X
	24	Sound and picture quality			X
	25	Video duration		X	
	26	Usability	X		
Personal-related quality criteria	27	Actor/actress			X
	28	Visibility of the explaining person		X	

General Limitations

Even after using the systematic assessment instrument, a certain degree of subjective interpretation always remains. In authentic classroom practice and lesson planning, the assessment instrument might take a lot of time or effort, or, if used by learners, require a certain amount of familiarisation. The study did not investigate the daily implementation for in-service teachers or the possible long-term effects of using the instrument, such as potential improvements in diagnostic competencies over time. In addition to the expert study, other validation methods, such as feedback from learners or direct measurements of learning outcomes after video use, have not been investigated. It is possible that not all criteria are equally relevant for different age groups or educational levels, as this study focused mainly on secondary education. Finally, this research and development of the instrument did not consider the rapid technological changes, such as AI-generated or interactive videos, which might have an influence on video creation and evaluation in the near future.

Theoretical Added Value

Beyond its practical utility, this study also provides a theoretical contribution by systematically integrating concepts from chemistry education research, multimedia learning theories, and instructional video assessment into a unified evaluation framework. While most previous approaches simply focused on one single analytical perspective on explanatory videos, the developed instrument aggregates, compares, and operationalizes different quality criteria specifically adapted to the context of chemistry, which was not exhaustively done before. Therefore, this research and the presented instrument combine multiple theoretical perspectives by bridging empirical findings from different research fields, e.g., formative assessment, learning with digital media, contextualization, or constructivism. For example, the possibility of an individual weighting of criteria acknowledges the importance of the context while learning. A theory-based assessment instrument for specific chemistry-educational media that synthesizes findings across these domains had previously been lacking. In addition, the presented instrument transfers abstract learning principles into practically applicable categories for easy video assessment.

Outlook

The presented tool has been used successfully and with positive feedback by students to create their chemistry explanatory videos. A guide for creating explanatory videos based on this instrument was created in a subsequent project. This is intended, for example, to instruct pre-service teachers of chemistry to create explanatory videos independently using the guide and then check the quality using the assessment instrument.

Further research could focus on criterion validity by further investigating the consistency of the experts' formulations with the assessment criteria and the systematic procedure for identifying empirically proven quality criteria. It would also be beneficial to intensively explore the links to the cognitive theory of multimedia learning (cf. Mayer & Chandler, 2001) and its correspondence with the assessment criteria. Additional research is required to ensure the comprehensibility and usability of the assessment instrument in comparison to traditional methods. The results show increased comprehensibility due to the more detailed criteria and characteristic information. However, the perspective of the video creators should also be considered in future studies to verify the comprehensibility from their point of view and to continuously develop the instrument further based on further evaluation studies. The authors also recommend further studies to verify the supposed positive effects on the diagnostic competence of users. Furthermore, investigations into how far the tool can support performance assessments would be of interest.

Conclusion

The developed assessment instrument represents a significant advance in evaluating chemistry explanatory videos. It offers a usable, differentiated, and empirically validated criteria list that is helpful in creating videos and selecting explanatory videos for chemistry lessons. The final 28 assessment criteria from different fields, based on a systematic literature search and an expert validation study, make the instrument universally applicable and allow a comprehensible assessment of the quality of instructional and explanatory videos.

References

Ahmad, Z., Ammar, M., Al-Thani, N. J., & Sellami, A. (2023). Effective Pedagogical Approaches Used in High School Chemistry Education: A Systematic Review and Meta-Analysis. *Journal of Chemical Education, 100*(5), 1796–1810. https://doi.org/10.1021/acs.jchemed.2c00739

Balcke, D. (2022). Erklärvideos – Eine kritische Analyse ihres Selbstanspruchs aus fach- und allgemeindidaktischer Perspektive. *Bildung und Erziehung*, *75*(1), 24-40. https://doi.org/10.13109/buer.2022.75.1.24

Beautemps, J., & Bresges, A. (2021). What Comprises a Successful Educational Science YouTube Video? A Five-Thousand User Survey on Viewing Behaviors and Self-Perceived Importance of Various Variables Controlled by Content Creators. *Frontiers in Communication*, *5*. https://doi.org/10.3389/fcomm.2020.600595

Bloom, B. S. (1984). Taxonomy of educational objectives. Boston: Allyn and Bacon.

Brass, M., Bekkering, H., Wohlschläger, A. & Prinz, W. (2000). Compatibility between Observed and Executed Finger Movements: Comparing Symbolic, Spatial, and Imitative Cues. *Brain and cognition*, *44*(2), 124-143. https://doi.org/10.1006/brcg.2000.1225

Chen, C.-M. & Wu, C.-H. (2015). Effects of different video lecture types on sustained attention, emotion, cognitive load, and learning performance. *Computers & Education*, *80*, 108-121. https://doi.org/10.1016/j.compedu.2014.08.015

Churches, A. (2008). *Bloom's digital taxonomy*. Available at: https://www.researchgate.net/publication/228381038_Bloom's_Digital_Taxonomy (Accessed 21.11.2023).

Coşkun, A., Büyükkeçeci, E., & Töre-Yargın, G. (2021). Exploring Success Criteria of Instructional Video Design in Online Learning Platforms. In Ö. D. Cordan, D. Arslan Dincay, Ç. Yurdakul Toker, E. Belkis Öksüz, & S. Semizoglu (Hrsg.) *Game + Design Education* (pp. 393-407). Springer. https://doi.org/10.1007/978-3-030-65060-5

Das Deutsche Schulportal (2022). *Erstmals repräsentative Daten zum Fernunterricht.* Availabel at: https://deutsches-schulportal.de/unterricht/lehrer-umfrage-deutsches-schulbarometer-spezial-corona-krise-april-2020/ (Accessed 04.12.2024).

de Jong, T., Linn, M.C., & Zacharia, Z.C. (2013). Physical and virtual laboratories in science and engineering education. *Science,* 340(6130), 305–308. https://doi.org/10.1126/science.1230579

Dorgerloh, S. & Wolf, K. D. (2020). *Lehren und Lernen mit Tutorials und Erklärvideos.* Beltz.

Findeisen, S., Horn, S. & Seifried, J. (2019). Lernen durch Videos: Empirische Befunde zur Gestaltung von Erklärvideos. *MedienPädagogik: Zeitschrift für Theorie und Praxis der Medienbildung*, 35, 16-36. https://doi.org/10.21240/mpaed/00/2019.10.01.X

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

Guo, P. J., Kim, J., & Rubin, R. (2014). How video production affects student engagement: An empirical study of MOOC videos. *Proceedings of the first ACM conference on Learning@ scale conference*, 41-50. https://doi.org/10.1145/2556325.2566239

Harmer, S. P. & Groß, K. (2021). CHEMideos – Fachdidaktische Analyse chemischer Erklärvideos. In S. Habig (Hrsg.). *Naturwissenschaftlicher Unterricht und Lehrerbildung im Umbruch?*. Gesellschaft für Didaktik der Chemie und Physik, Online-Jahrestagung, (pp. 589-592).

Harmer, S. P. & Groß, K. (2022). Developing and evaluating guidelines for analysing video tutorials for the chemistry classroom. In G.S. Carvalho, A.S. Afonso & Z. Anastácio (Eds.), *Fostering scientific citizenship in an uncertain world (Proceedings of ESERA 2021),* pp. 332-342, Braga, Portugal.

Heidig, S., Müller, J. & Reichelt, M. (2015). Emotional design in multimedia learning: Differentiation on relevant design features and their effects on emotions and learning. *Computers in Human Behavior, 44,* 81-95. https://doi.org/10.1016/j.chb.2014.11.009

Hommel, B. & Stränger, J. (1994). Wahrnehmung von Bewegung und Handlung. *Enzyklopädie der Psychologie, Themenbereich C Theorie und Forschung, Serie II Kognition, 1,* 529-603.

Honkomp-Wilkens, V., Wolf, K. D., Jung, P. & Altmaier, N. (2022). Informelles Lernen auf YouTube: Entwicklung eines Analyseinstruments zur Untersuchung didaktischer und gestalterischer Aspekte von Erklärvideos und Tutorials. *MedienPädagogik,* 18 (Jahrbuch Medienpädagogik), 495-528. https://doi.org/10.21240/mpaed/jb18/2022.03.08.X

Jegstad, K. M. (2023). Inquiry-based chemistry education: a systematic review. *Studies in Science Education, 60*(2), 251–313. https://doi.org/10.1080/03057267.2023.2248436

Jebe, F., Konietzko, S., Lichtschlag, M., & Liebau, E. (2019). *Studie: "Jugend/YouTube/Kulturelle Bildung. Horizont 2019".* Essen: Rat für Kulturelle Bildung e. V. Available at: https://www.bosch-stiftung.de/sites/default/files/publications/pdf/2019-06/Studie_Jugend%20Youtube%20Kulturelle%20Bildung%202019.pdf (Accessed 21.11.2023).

Korossy, K. (2022). *Modellierung schulischer Notenbildung.* Springer. https://doi.org/10.1007/978-3-658-36197-6

Krammer, K. & Reusser, K. (2005). Unterrichtsvideos als Medium der Aus- und Weiterbildung von Lehrpersonen. *Beiträge zur Lehrerbildung*, *23*(1), 35-50. https://doi.org/10.25656/01:13561

Kulgemeyer, C. (2018). Qualitätskriterien zur Gestaltung naturwissenschaftlicher Erklärvideos. In C. Maurer (Hrsg.), *Naturwissenschaftliche Bildung als Grundlage für berufliche und gesellschaftliche Teilhabe.* Gesellschaft für Didaktik der Chemie und Physik, Jahrestagung in Kiel. (pp. 285-288).

Liu, C., Chen, S., Liu, S., Liu, J., Wang, J., & Han, Y. (2021). A Delphi Study and Verification of Assessment of Chemistry Experiments Videos. *Croatian Journal of Education, 23*(3), 903-934. https://doi.org/10.15516/cje.v23i3.4018

Marquardt, K. (2016). *Beurteilungsraster für Mathematik-Erklärvideos: Chancen, Grenzen und Durchführung einer Operationalisierung mittels Resultaten aus der Schulbuchforschung*. [Diplomarbeit] Universität Wien. https://doi.org/10.25365/thesis.43166

Martin, J.-P. & Kelchner, R. (1998). Lernen durch Lehren. In J. P. Timm (Hrsg.), *Englisch lernen und lehren - Didaktik des Englischunterrichts*. (pp. 211-219). Cornelsen.

Mayer, R. (2014). *The Cambridge Handbook of Multimedia Learning (2nd ed.)*. Cambridge University Press. https://doi.org/10.1017/CBO9781139547369

Mayer, R. & Chandler, P. (2001). When learning is just a click away: Does simple user interaction foster deeper understanding of multimedia messages? *Journal of Educational Psychology, 93*(2), 390-397. https://psycnet.apa.org/doi/10.1037/0022-0663.93.2.390

Meller, S. (2017). Erklärvideos als Bildungsressource – Wie YouTube und Co. Lernen verändern. *Ph publico, 12*, 115-124, https://doi.org/10.25656/01:28383

Moher, D., Liberati, A., Tetzlaff, J. & Altman, D. (2011). Bevorzugte Report Items für systematische Übersichten und Meta-Analysen: Das PRISMA-Statement. *DMW-Deutsche Medizinische Wochenschrift, 136*(08), 9-15. https://doi.org/10.1055/s-0031-1272978

Müller, F. & Oeste-Reiß, S. (2019). Entwicklung eines Bewertungsinstruments zur Qualität von Lernmaterial am Beispiel des Erklärvideos. In J. M. Leimeister & K. David (Hrsg.), *Chancen und Herausforderungen des digitalen Lernens.* (pp. 51-73). Springer. https://doi.org/10.1007/978-3-662-59390-5_4

Naggar, L. (2022). *Erklärvideos im Mathematikunterricht in Grundschulen–Eine mehrperspektivische und qualitative Untersuchung zu möglichen Qualitätsmerkmalen von Erklärvideos*. [Masterarbeit] Freie Universität Berlin. Refubium – Repositorium der freien Universität Berlin. https://refubium.fu-berlin.de/handle/fub188/33798

Nerdel, C. (2017). *Grundlagen der Naturwissenschaftsdidaktik: Kompetenzorientiert und aufgabenbasiert für Schule und Hochschule*. Berlin, Heidelberg: Springer Spektrum. http://doi.org/10.1007/978-3-662-53158-7

Nordhausen, T. & Hirt, J. (2020). RefHunter. Manual zur Literaturrecherche in Fachdatenbanken. Version 5.0 (Martin-Luther-Universität Halle-Wittenberg & OST (ehemals FHS St.Gallen), Hrsg.). Halle (Saale). Avaiable at: https://refhunter.eu/manual/ (Accessed 21.11.2023).

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimhaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., McGuinness, L. A., Stewart, L. A., Thomas, J., Tricco, A. C., Welch, V. A., Whiting, P., & Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ (Clinical research ed.)*, 372(71). https://doi.org/10.1136/bmj.n71

Plass, J. L., Heidig, S., Hayward, E. O., Homer, B. D. & Um, E. (2014). Emotional design in multimedia learning: Effects of shape and color on affect and learning. *Learning and Instruction*, *29*, 128-140. https://doi.org/10.1016/j.learninstruc.2013.02.006

Preston, G., Phillips, R., Gosper, M., McNeill, M., Woo, K., & Green, D. (2010). Web-based lecture technologies: Highlighting the changing nature of teaching and learning. *Australasian Journal of Educational Technology*, *26*(6), 717-728. https://doi.org/10.14742/ajet.1038

Renkl, A. (1997). *Lernen durch Lehren*. Springer. https://doi.org/10.1007/978-3-663-08696-3

Riedl, A. (2008). Innere Differenzierung–Herausforderung für modernen Unterricht. In Föderale Ausbildungsagentur, Staatliche technische Uraler Universität (Hrsg.), *Wirtschaft und Linguistik: Wege einer Wechselwirkung. Eine Sammlung von Materialien einer inernationalen wissenschaftlich-praktischen Konferenz von Studierenden und Doktoranden* (pp. 122-128). Jekaterinburg: USTU-UPI. https://www.yumpu.com/de/document/view/27692177/innere-differenzierung-lehrstuhl-fur-padagogik-technische-

Rummler, K. & Wolf, K. D. (2012). Lernen mit geteilten Videos: aktuelle Ergebnisse zur Nutzung, Produktion und Publikation von Onlinevideos durch Jugendliche. In W. Sützl, F. Stalder, R. Maier, & T. Hug (Hrsg.), *Kulturen und Ethiken des Teilens* (pp. 253-266). Innsbruck University Press. https://doi.org/10.25969/mediarep/2163

Rüsseler, M., Sterz, J., Kalozoumi-Paisi, P., Schill, A., Bender, B., Hoefer, S. H., Schleicher, I., Damankis, A. I., Josephs, D., Ochsendorf, F., Stefanescu, C., Hoffmann, H., Schreckenbach, T. & Walcher, F. (2017). Qualitätssicherung in der Lehre–Entwicklung und Analyse von Checklisten zur Beurteilung von Lehrvideos zum Erlernen praktischer Fertigkeiten. *Zentralblatt für Chirurgie-Zeitschrift für Allgemeine, Viszeral-, Thorax- und Gefäßchirurgie*, *142*(01), 32-38. https://doi.org/10.1055/s-0042-124416

Seethaler, S., Burgasser, A. J., Bussey, T. J., Eggers, J., Lo, S. M., Rabin, J. M., Stevens, L. & Weizman, H. (2020). A research-based checklist for development and critique of STEM instructional videos. *Journal of College Science Teaching*, *50*(1), 21-27. http://dx.doi.org/10.1080/0047231X.2020.12290671

Sommer, K., Wambach-Laicher, J., & Pfeifer, P. (2018). *Konkrete Fachdidaktik Chemie - Grundlagen für das Lernen und Lehren im Chemieunterricht*. Seelze: Friedrich Verlag GmbH.

Sterzing, F., Szabone Varnai, A., & Reinhold, P. (2021). Theoriebasierte Gestaltung & Validierung der Qualität von Erklärvideos. In S. Habig (Hrsg.), *Naturwissenschaftlicher Unterricht und Lehrerbildung im Umbruch?*. Gesellschaft für Didaktik der Chemie und Physik, Online Jahrestagung 2020. (pp. 113-117).

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

Tenberg, R. (2021). *Didaktische Erklärvideos: Ein Praxis-Handbuch.* Franz Steiner Verlag.

Wolf, K. D. (2015). Video-Tutorials und Erklärvideos als Gegenstand, Methode und Ziel der Medien- und Filmbildung. In A. Hartung, T. Ballhausen, C. Trültzsch-Wijnen, A. Barberi, & K. Kaiser-Müller (Hrsg.), *Filmbildung im Wandel* (pp. 121-131). New Academic Press.

Appendix

- Description of the assessment criteria, including empirical evidence (Table S1)
- Complete overview of the expert statements and criteria in the video evaluations and comparison with the criteria of the assessment instrument (Table S2)
- Assessment instrument as an Excel spreadsheet with automated calculation of the overall score

Table S1

Description of the assessment criteria, including empirical evidence

	Quality criteria	**Description/Explanation**	**Empirical evidence**
Science education and content-related quality criteria			
1	Correctness	The subject content is presented correctly. Statements are explained comprehensively. Scientific sources are used and cited.	(Beautemps & Bresges, 2021; Naggar, 2022)
2	Completeness	The content is complete (including experiences and various solutions or alternative notations).	(Marquardt, 2016; Müller & Oeste-Reiß, 2019; Naggar, 2022; Sterzing, Szabone Varnai, & Reinhold, 2021)
3	Precision	The content is precise, i.e. the focus is on the core contents and unnecessary excursions, or additional information are avoided. There are exact statistical values and numbers.	(Marquardt, 2016; Müller & Oeste-Reiß, 2019; Naggar, 2022; Sterzing et al., 2021)
4	Complexity and scope	The complexity and the scope of the content are appropriate.	(Müller & Oeste-Reiß, 2019; Naggar, 2022)
5	Didactic reduction and target group orientation	The content is adapted to preknowledge, misconceptions and interest of the learners. The degree of mathematics is taken into account.	(Kulgemeyer, 2018; Marquardt, 2016; Naggar, 2022; Sterzing et al., 2021)
6	Attractiveness of the content	The content of the video appears to be attractive for the viewer, as it … a) … is current b) … is informative c) … is relevant for the target group (e.g., everyday knowledge, knowledge for a test) d) … is explaining a new concept e) … is embeded into a greater context	(Harmer & Groß, 2021; 2022; Kulgemeyer, 2018; Müller & Oeste-Reiß, 2019; Naggar, 2022; Sterzing et al., 2021)

		f) … is interdisciplinary, i.e., it shows understanding of application and transfer possibilities to other disciplines g) … shows a relation to everyday life or professional practice (concrete application of the contents)	
Structural quality criteria			
7	Comprehensibility and coherence in the structure	The structure of the presented content in the video is comprehensible and understandable. The explanation video ... 1. ... starts with an introduction to the topic (e.g., stating the question) 2. ... then mentions the concrete topic of the video. 3. ... offers a differentiated table of contents to an adequate extent. 4. ... explains the content in a structured way (e.g., with increasing difficulty or a slight arc of suspense) 5. ... closes with a summary of the most important results/contents.	(Beautemps & Bresges, 2021; Kulgemeyer, 2018; Marquardt, 2016; Müller & Oeste-Reiß, 2019; Naggar, 2022)
8	Separation of structural elements & segmentation & cross-connections	There is a clear differentiation between definitions, statements, proofs, examples, tasks and other structural elements. If the content is complex, it is divided into various parts. There are explicit hints towards cross connections.	(Marquardt, 2016; Sterzing et al., 2021)
9	Design structure/ emphasis principle	The structure is supported by several design elements. Important information is highlighted so that the organization of the learning content gets clear.	(Marquardt, 2016; Sterzing et al., 2021)
10	Clarity of the explanatory structure	The structure of the explanations is logical and consistent. Thoughts are structured as linearly and transparently as possible. If specialist knowledge is the learning goal, a rule-example structure is preferred; for routines, an example-rule-explanation structure is preferred.	(Kulgemeyer, 2018; Müller & Oeste-Reiß, 2019; Sterzing et al., 2021)
Science education and methodological quality criteria			
11	Appropriate use of illustration tools	Examples, analogies and models are appropriately introduced and used in the video.	(Kulgemeyer, 2018; Sterzing et al., 2021)
12	Adequate use of experiments	The experiment … a) … supports the communication of the content. b) … supports a research-based approach. c) … is novel and interesting.	(Liu et al., 2021)
13	Adequate implementation of experiments	a) The methods and steps of the experiment are clear and comprehensible. b) Instruments and chemicals are neatly on the experiment table. c) The experimental environment is quiet.	(Kulgemeyer, 2018; Liu et al., 2021; Sterzing et al., 2021)

		d) The experimental device is used correctly. e) The amount of chemicals used is appropriate and there is no waste. f) Necessary safety aspects were taken into account.	
14	Use of differentiation measures	The video meets the demands of (internal) differentiation. This was implemented using: a) Notes, additional material, in-depth material, additional offers b) Exercises, subsequent learning tasks, tests, quiz questions c) References to further media and literature d) Alternative and differently complex solutions e) Consideration of different learning styles f) Playback control and/or pausing option	(Delen, Liew, & Willson, 2014; Hasler, Kersten, & Sweller, 2007; Kulgemeyer, 2018; Marquardt, 2016; Mayer & Chandler, 2001; Naggar, 2022; Schwan & Riempp, 2004; Spanjers, Van Gog, Wouters, & Van Merriënboer, 2012; Zahn, Barquero, & Schwan, 2004; Zhang, Zhou, Briggs, & Nunamaker Jr, 2006)
15	Target group orientation and comprehensibility of the language	An addressee-oriented and understandable language is used. This is shown in: a) A simple language style with clear sentence structure b) Avoiding redundant words c) Brevity and conciseness d) Choice of words appropriate to the target group e) Specialist language tailored to the target group	(Coşkun, Büyükkeçeci, & Töre-Yargın, 2021; Kulgemeyer, 2018; Marquardt, 2016; Müller & Oeste-Reiß, 2019; Naggar, 2022; Sterzing et al., 2021)
16	Action and activity orientation	This includes: a) Operational principle, i.e. the learning objects are developed in an action-oriented manner by building a system of operations b) Promoting active learning c) Activity orientation stimulates reaction or arouses curiosity by presenting situations and problems. d) Invitational character through a direct and personal address → Scientific skills and problem-solving skills are promoted	(Fiorella, van Gog, Hoogerheide, & Mayer, 2017; Marquardt, 2016; Müller & Oeste-Reiß, 2019; Naggar, 2022; Sterzing et al., 2021)
Science education and pedagogical quality criteria			
17	Learning objectives and curriculum	Sufficient reference is made to the skills and knowledge the learners should strive for. Learning objectives are formulated and reasons are given as to why it is necessary and useful to learn certain facts (legitimization of learning objectives). The selected content supports these learning objectives or the curriculum.	(Marquardt, 2016; Müller & Oeste-Reiß, 2019)

18	Increasing motivation and attention	The teaching/explanatory video arouses interest and promotes the target group's attention and motivation.	(Müller & Oeste-Reiß, 2019)
19	Emotional sensitivity & demanding a serious attitude	The explanatory video triggers emotional feelings (funny, exciting, shocking...) and still demands serious engagement with the content.	(Marquardt, 2016; Naggar, 2022)
20	Transferability to different learning environments	The video can be used in various learning environments. This enables independent learning.	(Müller & Oeste-Reiß, 2019)
21	Neutrality	The content is neutral regarding gender, religion, nationality, etc. The video does not contain any discrimination, degradation, prejudice or one-sided role models. Also: Taking into account possible color vision deficiency: The colors red and green are not used to differentiate.	(Marquardt, 2016; Müller & Oeste-Reiß, 2019)
Media design and technical quality criteria			
22	Aesthetics and design	The design and layout appear aesthetic, creative and appealing. The design is logical and consistent. Recurring symbols, elements or structures are used.	(Heidig, Müller, & Reichelt, 2015; Müller & Oeste-Reiß, 2019; Naggar, 2022; Plass, Heidig, Hayward, Homer, & Um, 2014; Sterzing et al., 2021)
23	Addressing different channels of perception	There are changes in presentation when conveying learning content.	(Marquardt, 2016; Naggar, 2022; Sterzing et al., 2021)
24	Sound and picture quality	The explanatory video has appropriate sound and image quality. This also includes: a) Clarity of writing and drawings b) Appropriate speed of animations	(Coşkun et al., 2021; Marquardt, 2016; Naggar, 2022)
25	Video duration	The duration of the video is appropriate. (Recommendation: 3-10 minutes)	(Beautemps & Bresges, 2021; Coşkun et al., 2021; Guo, Kim, & Rubin, 2014; Müller & Oeste-Reiß, 2019; Naggar, 2022)
26	Useability	The video is characterized by the following attributes: • User-friendliness • Control option/navigation option • Longevity • Accessibility/availability • Reasonable costs • Freedom of Advertisements	(Müller & Oeste-Reiß, 2019) (Naggar, 2022)
Personal-related quality criteria			
27	Actor/actress	1. The actor creates trust through a short introduction at the beginning, if necessary with an academic degree/title. Justification why the actor is an expert.	(Beautemps & Bresges, 2021; Coşkun et al., 2021; Müller & Oeste-Reiß, 2019; Naggar, 2022)

		2. The actor uses an appropriate expression and tone of voice. This supports attentive listening. 3. The personality of the actor is appealing (authentic, committed, fun, likeable, humorous, friendly, happy, open charisma).	
28	Visibility of the explaining person	The visibility of the person explaining is appropriate for the video. The face, demeanor and external appearance of the person explaining the video does not distract from the content.	(Chen & Wu, 2015; Guo et al., 2014; Kizilcec, Bailenson, & Gomez, 2015; Ouwehand, van Gog, & Paas, 2015; Van Gog, Verveer, & Verveer, 2014)

Table S2

Complete overview of the expert statements and criteria in the video evaluations and comparison with the criteria of the assessment instrument

Nr.	Main Categories	Mentions	Exp. 1	Exp. 2	Exp. 3	Assess-ment	Mentions per criterion
1	**Correctness**	"Technical correctness"	7	14	4		25
2	**Completeness**	"Technical terms are included"	1				14
		"Lack of relevant technical content, reaction equations or sub-steps"	5	2	2		
		„Explanations are missing"			1		
		"Technical content is not conveyed either"	1				
		"Shortened very strongly"	1				
		"Not all molecule names and substance groups are introduced."	1				
3	**Precision**	"Technical language is imprecise"	1				4
		"Concise"			2		
		"Strong simplification"	1				
4	**Complexity and scope**	"Too much information"	1				5
		"Unnecessary additional information"			3		
		"Difficult to focus"	1				
5	**Didactic reduction and target group orientation**	"Different symbols on different levels"	1				7
		"Too much knowledge is assumed"	2		2		
		"Reduced by too much"	1				

		“Symbols with different levels of abstraction”	1				
6	**Attractiveness of the content**	“Everyday phenomenon“	1	4	3		8
7	**Comprehensibility and coherence in the structure**	“Builds up ... well”	1	1			46
		“Introduction / instructions at the beginning”			9		
		“Objective“		3			
		“Structure is described“			1		
		“Didactically building on one another”			1		
		“Topic is mentioned at the beginning”		7	2		
		“Understandable”	2				
		“Presentation of the most important terms”			1		
		“It is absolutely not clear to me what this is about.”	1				
		“Clarity and formulation of the objective“	3				
		“Very difficult to follow”	1				
		“Lacks any kind of structure”	1				
		“Summary at the end”	1	2	9		
8	**Separation of structural elements & segmentation & cross-connections**	“Phasing of the process”	1				10
		“Well structured”			1		
		“Zoomed in on key points so that it is always clear which part of the process is involved”	1	1			
		“Introduction missing”	1				
		“Back connections (precise and comprehensible)”	3				
		“Connections”	2				
9	**Design structure / emphasis principle**	“Central elements ... highlighted (highlighting principle)”	5				5
10	**Clarity of the explanatory structure**	“Not explained why …”	2	2	2		15
		“Cognitive load theory considered“			1		
		“Explained in a comprehensible way”		5			
		“Explanations“	1	2			
11	**Appropriate use of illustration tools**	“Reference to everyday life”	1				20
		“Nice, but cognitive load too high”			2		
		“Appropriate use“		1			
		“Few illustrations”		1			
		“Labels missing”			1		

		„Use of models“	2	2	1		
		“Very messy and not up to the standard”	1				
		“Everyday relevance”	2	1	1		
		“Analogues are used”	3		1		
12	**Adequate use of experiments**	“Why is the experiment even being carried out?”	1				5
		“Selection of chemicals“	1	1	2		
13	**Adequate implementation of experiments**	“Procedure / conduct of the experiment”	3	5	3		57
		“Experiment is not carried out, only explained”			1		
		“Experimental: moderate”			1		
		“Use of chemicals“		3	1		
		“Basics of safe experimentation”	3	2			
		“Nicely implemented” / “Not so nice”			2		
		“Environment tidy” / “Unnecessary materials“			1		
		“Messy work”			1		
		“Why aren't the materials ready already?”			1		
		“Design rules followed”			2		
		“Structuring”	1	3			
		“Experiment observation”	2				
		“Effect difficult to recognize”			1		
		“Clearly visible“	2	4	3		
		“Description of the experimental setup”	4	2	1		
		“Label missing”	1	1	2		
		“Result is anticipated”	1				
14	**Use of differentiation measures**	“Explanations are brief, reference to other video”		1			1
15	**Target group orientation and comprehensibility of the language**	“Appropriate technical language”		10	1		17
		“Appropriate language”		3	3		
16	**Action and activity orientation**	“Result is anticipated”	2	2	1		7
		“Joint creation of the reaction equation”			1		
		“Encourages thinking along“		1			
17	**Learning objectives and curriculum**						0

18	**Increasing motivation and attention**	“Motivating introduction“		15	7		22
19	**Emotional sensitivity & demanding a serious attitude**						0
20	**Transferability to different learning environments**						0
21	**Neutrality**						0
22	**Aesthetics and design**	“Nicely designed”, “Illustrative”, “Appealing”	2	6	3		40
		“Uniform design / fonts“		5			
		“Blackboard is not clean”	1	1			
		“Misleading representations“			2		
		“Consideration principles according to Mayer (2014)”			2		
		“Low cognitive load in the representations”			2		
		“Representations do not distract“	1				
		“Drawings, illustrations, graphics“			8		
		“Transitions“			1		
		“Changes in representation“	1				
		“Selected representations encourage misconceptions“	1				
		“Image is too full”	1	1			
		“Arrows for visualization“	1				
		“Empty slides for long durations are irritating”	1				
23	**Addressing different channels of perception**	“Correspondence between spoken / text / image”	3	3	4		17
		“Relation between text, image, spoken	1		2		
		“Time delay between illustrations and spoken text”	2		1		
		“Technical terms not … noted in the appropriate places”	1				
24	**Sound and picture quality**	“Picture quality“	2	14	7		58
		“Sound quality“	1	14	9		
		“Not legible / recognizable“	3	7	1		
25	**Video duration**	“Longer processes are not sped up”	1				4
		“Short and concise”			1		
		“A little too slow” / “Very slow”			2		
26	**Usability**						0

27	**Actor / actress**	“Accent“	1				16
		“Effect of the voice and emphasis”	4	2			
		“Unprofessional formulations“			1		
		“Way of speaking“	2	6			
28	**Visibility of the explaining person**	“Speaker's clothing appears unprofessional“		1			3
		“Involvement of the person“			2		
29	**Additional Criteria**	“Background music / noises“	3	6	6		15
		“Chain reaction hardly clear”	1				1
		“Everything clearly visible“	2	5	1		8
		“Abstraction of the matter”	1				1
		“Not an explanatory video” / “More of an experimental video”	1	1	2		4
		“Didactically well presented“			1		1
		“Often a bit complicated”			1		1
		“Technically elegant” / “Technically well implemented” / “Technically OK, but not exactly super good” / “Technically: bad”			7		7
		“Didactically completely pointless” / “Didactically completely unnecessary” / “Didactically: OK” / “Didactically bad”			3		3
		“Strange transition” / “Strange transitions”			2		2
		“Works from right to left”			1		1
		Nominations per expert:	120	173	157		450

Category weighting	**Color**
Very important	
Important	
Not important	

Assessment instrument as an Excel spreadsheet with automated calculation of the overall score

The assessment instrument can be found in a separate Excel file.

Corresponding Author Contact Information:

Author name: Prof. Dr. Jenna Koenen

Department: TUM School of Social Sciences and Technology, Professorship of Chemistry Education

University, Country: Technical University of Munich, Germany

Email: jenna.koenen@tum.de

Please Cite: Merk, S., Diermann, D., Forster, K., & Koenen, J. (2025). A criteria-based assessment instrument to assess the quality of explanatory videos in chemistry education. *Journal of Research in Science, Mathematics and Technology Education*, *8*(SI), 161-194. DOI: https://doi.org/10.31756/jrsmte.417SI

Conflict of Interest: The authors have no conflicts of interest to declare.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Data Availability Statement: The research data is stored on a server at the Technical University of Munich. It can be made available upon request via email to Prof. Dr. Jenna Koenen at jenna.koenen@tum.de.

Ethics Statement: All research was conducted in accordance with ethical guidelines, data security considerations, and the journals policies. The declarations of consent have been received.

Author Contributions: The article was based on Ms Merk's master's thesis. She was therefore primarily responsible for the literature research, data acquisition and analysis, and the initial manuscript draft. The other three authors supervised the work, providing feedback, edited, and finalized the article in line with the revisions.

Received: February 14, 2025 ▪ Accepted: May 20, 2025