Expert Reasoning in Fragment-Level Quality Metadata Authoring for Learning-Sensitive Educational Video Adaptation

Syed Asim Jalal[1]*, Ayesha Khattak[2], Nicholas Gibbins[3], David Millard[3], Shaukat Ali[1], Muhammad Haseeb[1],
[1]Department of Computer Science, University of Peshawar
[2]School of Computer Science and IT, Institute of Management Sciences, Peshawar, Pakistan
[3]School of Electronics and Computer Science, University of Southampton, United Kingdom
*Correspondence: sa.jalal@uop.edu.pk

Citation | Jalal. S. A, Khattak. A, Gibbins. N, Millard. D, Ali. S, Haseeb. M, "Expert Reasoning in Fragment-Level Quality Metadata Authoring for Learning-Sensitive Educational Video Adaptation", IJIST, Vol. 8 Issue. 6 pp 2433-2451, September 2026
DOI: https://doi.org/10.33411/IJIST/2012
Received | Aug 16, 2026 **Revised |** Sep 01, 2026 **Accepted |** Sep 05, 2026 **Published |** Sep 08, 2026.

Limited battery power in mobile devices is an important concern in mobile learning activities based on educational multimedia. Power-saving multimedia adaptation approaches extend battery life by reducing the quality of streamed multimedia. However, these approaches often do not consider the sensitivity of visual instructional content to quality reduction. In learning-sensitive educational video adaptation, visual quality reduction must be guided by the instructional importance in each video segment. This requires fragment-level metadata that identifies segment boundaries and specifies the minimum acceptable quality required to preserve perceived visual comprehensibility. Producing such metadata requires judgments about instructional meaning, visual clarity, and acceptable quality thresholds. This paper reports an exploratory study of expert reasoning during manual segment-level quality metadata authoring. Five experts at the University of Southampton, UK, reviewed an educational video in four presentation qualities. Their task was to identify fragment boundaries and assign the minimum acceptable quality (MAQ) required to preserve the perceived comprehensibility of instructional information. Think-aloud data and post-task questionnaire responses were analysed to identify decision cues and authoring strategies. Results indicated that the metadata generation task was perceived as clear, manageable, and feasible. Descriptively, experts reported high task clarity (M = 4.6), confidence in MAQ assignment (M = 4.2), and manageability and feasibility (M = 4.6), while ratings for mental effort and difficulty in judging visual sufficiency were lower (M = 2.8 and M = 2.4, respectively). Experts were able to identify twelve fragments, while MAQ choices showed some variation across experts. Higher quality levels were generally selected for fragments containing important text or diagrams, whereas lower quality or audio-only presentation was considered acceptable when the information was mainly carried by narration. The contribution of this study is to characterize the human reasoning involved in fragment-level quality metadata authoring and to derive implications for future semi-automated authoring tools.

Keywords: Mobile Learning; Human-in-the-Loop Annotation; Metadata Authoring; Educational Video; Learning-Sensitive Adaptation; Content-Aware Adaptation

Introduction:

Mobile learning, which means engaging in learning activities and/or learning skills using portable digital devices such as smartphones and tablets, has grown quickly over the last decade [1]. This growth is due to the increased affordability of high-performance mobile devices and the availability of high-data-rate wireless connectivity [2]. Improvements in mobile technologies, such as screens, processors, and higher-bandwidth internet connectivity, enable mobile users to stream educational videos at higher qualities at any time and from any place [3]. This ubiquitous access has expanded learning opportunities. Open Educational Resources (OERs) based on rich instructional media are now freely available for both formal and informal learning. Video-based learning has become increasingly important in both formal and informal educational settings [4]. Recent research on video lectures also shows that visual cues can influence learner attention and retention, further highlighting the importance of preserving instructionally relevant visual information during video delivery [5].

Mobile learning, however, still has some technical constraints. Limited battery power on mobile devices is one major constraint, as battery technology has not seen improvements at the same rate as other hardware and network connectivity technologies. Streaming multimedia over wireless networks consumes significant battery power [6]. Research suggests that streaming video through a wireless network can use nearly twice as much battery power compared to playing the same video stored locally on the device [7]. Many studies have looked at different personalization and adaptation issues in mobile learning, based on device differences, network characteristics as well as learner needs [8]. Adaptation targeting limited battery life remains comparatively understudied in the context of mobile learning. Some researchers have suggested energy-saving features to be incorporated into mobile learning platforms [9].

One general approach used in reducing battery power consumption during multimedia streaming is to reduce the volume of data streamed by lowering the visual quality of the multimedia content [6][7][10][11][12][13][14]. High-quality videos require more data transfer and need more processing power. By adopting the strategy of reducing encoding parameters like resolution, frame rate etc. and compression settings, power consumption during multimedia streaming can be reduced on mobile devices. This means the device receives less data over the wireless network. However, this approach introduces a compromise between power-preservation and the quality of experience. In the context of educational multimedia, this compromise may be consequential. Lowering visual quality beyond a certain threshold may hinder comprehension of some visual instructional information. This may be more evident in videos that include detailed graphics or textual contents in smaller font sizes.

Many energy-saving techniques have relied on uniformly lowering the visual quality of an entire video stream, with little consideration for the impact on the instructional value [6][7][11][12]. This one-size-fits-all approach is often used as a power-saving-motivated strategy, but may leave some visual content difficult to understand. Uniform degradation of educational multimedia may negatively affect the perceived visibility and comprehension of visual learning contents, such as on-screen text, diagrams, or highlighted examples. This may ultimately impair learning experience. Our earlier work empirically confirmed this concern, showing that indiscriminate quality reduction can reduce the perceived clarity of visually dependent instructional content [15].

Some researchers have explored generic content-aware and adaptive streaming methods in order to identify video segments where lower quality may be less likely to affect the viewer's experience. These approaches selectively reduce quality [16][17][18]. Some content-aware adaptation emphasizes the importance of fine-grained control at the fragment level during adaptation. For example, deep learning models are now capable of detecting perceptually salient scenes in real time, based on which they enable adaptive bitrate allocation

across different segments [19][20][21]. Machine learning has been applied to identify educationally significant segments by analyzing multimodal cues [22][23].

Despite advancements in multimedia adaptation techniques, approaches that incorporate instructional considerations into their adaptation decisions are still lacking. To address these limitations of generic quality-reduction based adaptation techniques in the context of educational multimedia adaptation, we proposed the Content-Aware Power-Saving Educational Multimedia Adaptation (CAPS-EMA) approach [24]. CAPS-EMA proposes a segment-level adaptation approach that lowers the quality of individual video fragments based on the visual instructional content. This approach avoids uniform degradation of visual quality. CAPS-EMA enables this based on fragment-specific metadata. This allows segments that are resilient to quality reduction to be streamed at lower quality than segments that are more sensitive to quality reduction. This offers a more balanced trade-off between battery conservation and perceived clarity of instructional visuals.

CAPS-EMA overcomes the challenge of visual information loss using a metadata-driven approach. Each portion of the multimedia that has similar visual details is associated with metadata. The metadata defines its minimum acceptable quality constraint. This is to ensure that learning visual information remains perceptible to learners after quality reduction. The constraint metadata is used by the streaming engine in real time, in order to stream each segment at the lowest visual quality that satisfies the minimum acceptable quality (MAQ) constraint defined by experts. Nevertheless, the CAPS-EMA framework does not specify any specific method to produce this fragment-level metadata. The fragment-level metadata annotation may be entirely automated, semi-automated with human oversight, or manual. Further research is required in this direction to identify effective and practical approaches for generating this metadata.

Manual metadata creation is not an unusual practice in multimedia systems. Instructional designers normally annotate educational resources using standard metadata models, such as Learning Object Metadata (LOM), specifically, when tagging with a variety of pedagogical attributes like learning outcomes, or content difficulty levels. In multimedia applications, human annotators often label key frames, scenes, or temporal segments for purposes such as video indexing, accessibility enhancement, or instructional highlighting. These precedents suggest that manual or expert-assisted metadata creation is a plausible and justifiable approach for systems like CAPS-EMA.

The unresolved problem addressed in this paper is not whether quality reduction can save battery power, but how the fragment-level metadata required for learning-sensitive adaptation can be authored. A system such as CAPS-EMA requires metadata that identifies fragment boundaries and minimum acceptable quality constraints. These judgments are not purely technical because they depend on whether visual information is instructionally important and whether it remains understandable after quality reduction. Therefore, the metadata authoring task is also a human-factors problem.

In this paper, we focus on the metadata authoring step needed for fragment-level minimum acceptable quality constraints in CAPS-EMA. The CAPS-EMA concept is power-saving-motivated, but this study does not measure device energy use, battery savings, or network usage. We also do not measure learning outcomes such as test performance or retention. Instead, we study expert judgments about whether the visual instructional information in a segment remains understandable under reduced quality, and how experts reason while producing this metadata.

The distinct contribution of this study is its empirical characterization of expert reasoning during fragment-level MAQ metadata authoring, rather than an evaluation of adaptation effectiveness or energy savings.

The main contributions of this paper are as follows:

It reports an exploratory think-aloud study of expert reasoning during manual segment-level quality metadata authoring for learning-sensitive educational video adaptation.

It identifies the visual and instructional cues used by experts when defining fragment boundaries and assigning minimum acceptable quality levels.

It provides descriptive feasibility evidence from expert post-task ratings on task clarity, effort, manageability, and willingness to repeat the authoring process.

It derives design implications for future semi-automated metadata authoring tools that can support, rather than fully replace, expert judgment.

This study is guided by the following research questions:

RQ1: How do experts identify fragment boundaries when authoring segment-level quality metadata for educational video adaptation?

RQ2: What visual and instructional cues do experts use when assigning minimum acceptable quality levels?

RQ3: How do experts perceive the clarity, effort, manageability, and feasibility of the manual metadata authoring task?

Related Work:

Energy-Efficient Multimedia Adaptation:

Mobile video streaming significantly drains battery due to continuous wireless data transmission and the high computational cost of decoding. To address this, a broad body of research has explored energy-efficient delivery techniques, primarily by reducing encoding parameters such as resolution, bitrate, and frame rate [7][9][10][11][12]. Prior studies report that lowering these parameters can reduce energy use, depending on device, codec, and network conditions. Prior work reports energy savings (e.g., up to 62.7% in specific experimental setups) when lowering quality parameters without necessarily compromising playback continuity [7][25].

Dynamic Adaptive Streaming over HTTP (DASH) [26][27] is a standard protocol for adaptive streaming applications. In DASH, videos are segmented into small chunks available at multiple quality levels. DASH then adapts the quality dynamically based on network conditions. This characteristic of DASH has been used by different approaches for improving energy efficiency in streaming [28][29][30][31][32]. Some studies use scene complexity models in the adaptation decisions as well [33]. Systems, such as GreenABR [34], improve energy-aware streaming by modeling segment-level power impact. Such approaches, however, optimize for Quality of Experience (QoE) in streaming using quality reduction-based content-generic adaptation strategies, without considering the instructional sensitivity of visual learning contents. We cite these results as background; our study does not measure energy use.

Content-Aware Adaptation Techniques:

Energy-efficient streaming methods have now evolved to content-aware techniques. These methods adapt video quality based on visual complexity or perceptual relevance. Some systems, for example, bitDetect [35] use objective video quality metrics to inform the bitrate reduction process based on scene characteristics that preserve user-perceived quality. [36] leveraged DASH to recommend quality based on predefined complexity scores, while [22] employed "scene interestingness" score for visual importance in order to guide quality adaptation. [37] propose an approach that dynamically computes spatial complexity score to optimize resolution for power saving. Similarly, SSIM-based segmentation techniques [18] are used to classify scenes as detailed or non-detailed, streaming only the detailed scenes at high quality. Recently machine-learning-based techniques have also emerged, for example, DQN-based bitrate adaptation [38] and JND-aware scene classification [39] that improve perceptual fidelity while reducing redundancy. More recent mobile-streaming studies have also combined content-aware or energy-aware information with adaptive bitrate decisions, including MEC-assisted content-aware optimization and joint bitrate and upsampling approaches [40]. These

approaches mainly optimize QoE, bandwidth use, or energy-related objectives, rather than the instructional importance of visual information.

Despite advances, most approaches remain content-generic and not ready for application on educational multimedia as they overlook the perceived clarity of instructional visuals, learning experience, and pedagogical relevance in the quality adaptation of visual learning information. Metrics like scene interestingness or complexity are not always indicators of instructional value. Important learning elements such as small-font text, diagrams, and equations may degrade under perceptual-only schemes. CAPS-EMA addresses these issues by proposing a learning-sensitive and fragment-level adaptation framework. CAPS-EMA proposes to assign quality constraints based on instructional sensitivity rather than perceptual scores.

Energy-Efficient Adaptation in Educational Multimedia

In the context of mobile learning, researchers have explored various strategies to extend learning continuity under mobile battery constraints. Moldovan et al. [41] were among the first to emphasize integrating battery-aware adaptation in mobile learning systems. They identified that lowering multimedia quality is a learner-friendly mechanism to save battery-power during learning activities. Moldovan et al. developed Eco-Learn [41], a metadata-driven platform that dynamically adjusts multimedia quality based on learner and device profiles. They experimentally showed that bitrate reduction could conserve battery-power consumption with minimal impact on learners' experience. Their evaluations were limited to brief 20-30 seconds clips from different learning videos. Similarly, MediaMTool [42] automated the creation and selection of multiple-quality versions for educational videos, achieving up to 17% battery savings during playback. However, both systems used subjective user experience for evaluation and did not study the loss of visual instructional details as a result of adaptation.

Other research contributions such as DQAMLearn [43] and studies by Molnar and Muntean [44][45] integrated QoE and device metrics in adaptive streaming using DASH. However, adaptation decisions in these works were based on content genre, not on visual or pedagogical relevance details. They found that different format, such as mini-lectures or screen recordings, respond differently to reductions in quality. Some formats were more vulnerable to instructional loss than others.

Overall, prior work mainly adapts multimedia using perceptual quality, content category, network, or device conditions. The present problem differs because it requires fragment-level judgments about whether visual instructional information remains understandable after quality reduction.

Metadata in Educational Multimedia:

In many multimedia learning systems, metadata has been used as a foundational mechanism for various purposes, such as, organizing, retrieving, and adapting content. In educational contexts, metadata is used to describe the pedagogical properties of learning objects, such as instructional goals, content type, topic complexity, and learner profile. One of the most widely used metadata standards is the IEEE Learning Object Metadata (LOM). LOM provides a schema for describing learning resources with the aim to ensure interoperability across platforms [46]. LOM includes descriptors such as educational level, learning resource type, interactivity type, and typical learning time. These are especially useful in personalized and adaptive learning systems. Metadata is also helpful in facilitating accessibility by providing alternate formats (e.g., captions, transcripts, or simplified content) based on learner needs [47]. Digital libraries and repositories also use metadata in managing educational content [48]. Most existing metadata schemas, however, are focused on educational attributes.

Human-in-the-Loop Video Metadata Annotation: Human-in-the-loop (HITL) metadata generation or annotation is useful in multimedia systems where automated methods are not

efficient, for example, in detecting pedagogical or context-sensitive features in instructional content [49]. During human-based annotation tasks, domain experts apply judgment and instructional insight in labeling video segments, ensuring that key educational elements are accurately recognized and retained. Abrar et al. proposed a framework for characterizing educational video segments using video content and learner attention information [50]. Recent reviews of human-in-the-loop AI in education further emphasize the continuing role of human input when educational interpretation and automated system decisions interact [51].

Expert-based annotation is widely used in lecture video indexing, summarization, accessibility, and recommendation. In educational settings, instructors and instructional designers label video content with topic boundaries, learning objectives, or key visual elements to support better navigation and learning personalization [52]. In the area of accessibility, HITL has been used for generating closed captions, audio descriptions, and scene-level metadata. These require human-level interpretation of instructional intent.

Despite the time and labor involved, human annotation is considered indispensable in domains where instructional accuracy and pedagogical factors are important. The CAPS-EMA approach, which does not explicitly recommend any specific manual or automated way of metadata generation, could benefit more from human-in-the-loop annotations as it could offer a feasible and cognitively aligned mechanism for segmenting content and assigning quality constraints at the fragment level. Human-in-the-loop annotation is particularly relevant when interpretation of instructional meaning is required. However, prior educational video annotation work has not examined how experts' reason about fragment boundaries and minimum acceptable quality constraints for learning-sensitive adaptation. This specific authoring problem is the focus of the present study.

Metadata Authoring Context: Fragment-Level Quality Metadata in CAPS-EMA:

This section briefly explains CAPS-EMA only to define the metadata authoring task examined in this study. The purpose of the present paper is not to evaluate CAPS-EMA as a streaming system, measure energy savings, or test learning outcomes. Instead, the focus is on the expert reasoning required to produce fragment-level quality metadata.

Energy-efficient adaptation of educational videos often lacks sensitivity to instructional content. To address this gap, we previously proposed the Content-Aware Power-Saving Educational Multimedia Adaptation (CAPS-EMA) framework. The design intent is to avoid delivering more visual quality than is needed to keep key instructional visuals understandable; this paper does not measure energy use. Unlike generic or perceptually optimized streaming techniques, CAPS-EMA prioritizes pedagogical importance over purely visual or aesthetic cues. Figure 1 illustrates the CAPS-EMA process.

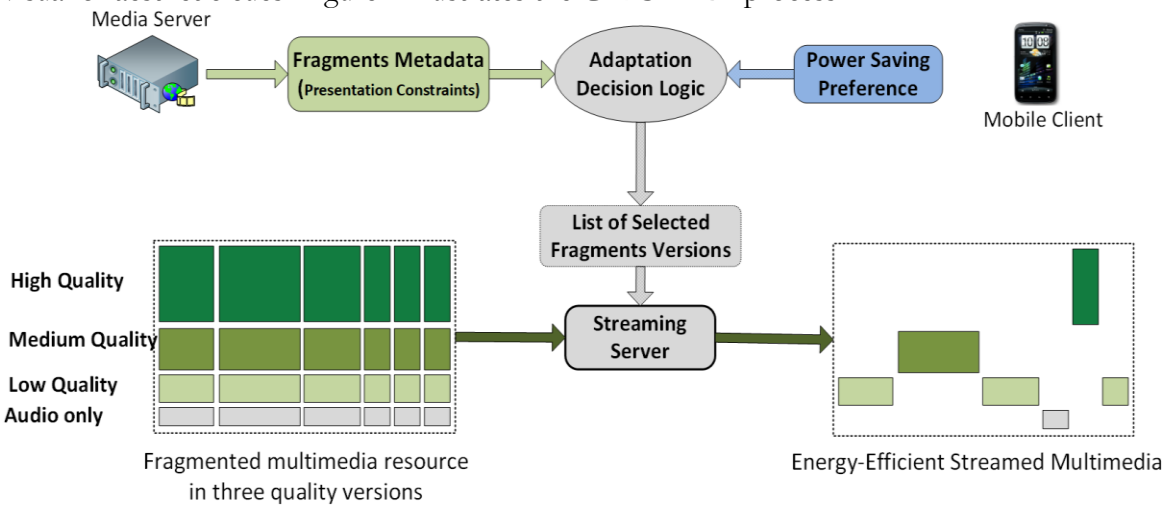

Figure 1. CAPS-EMA Process

At the heart of CAPS-EMA lies a fragment-level metadata-driven approach. Each educational video is segmented into instructional fragments. Fragments are temporally bounded segments that require distinct quality thresholds. A minimum acceptable quality constraint is assigned to each fragment. This quality-constraint may be determined through manual inspection or automated methods. These fragment-constraints are represented in metadata. The metadata then directs the system to deliver each fragment at its pedagogically required quality constraint. Fragments with crucial visuals (e.g., diagrams, equations) will have higher quality constraint. Visual information that is not quality-sensitive or narration-only segments may be streamed in lower quality.

This adaptation mechanism supports delivery at multiple quality levels. The fragment-level constraints help maintain the visibility of important instructional visuals. CAPS-EMA has some similarities in structure with common streaming protocols like DASH and HLS. Both of these protocols divide videos into segments for adaptive streaming. However, unlike these bandwidth-responsive streaming approaches, CAPS-EMA uses semantic and instructional metadata. This makes the approach more suitable for educational multimedia.

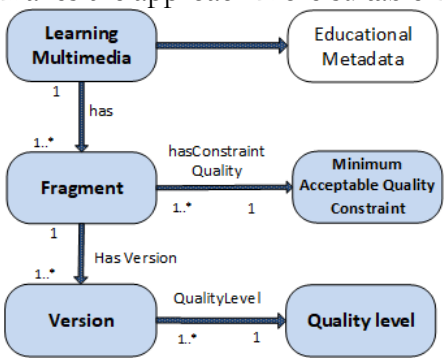

Figure 2. Conceptual Educational-Multimedia Metadata Model

The Fragmented Educational Multimedia Model (FEMM), shown in Figure 2, provides the basic structure for generating this metadata. FEMM describes how an educational multimedia resource is organized in a hierarchical way. It divides the multimedia into smaller fragments and their different versions. Each multimedia resource is split into one or more instructional fragments. Each fragment can have multiple versions, which may differ in encoding settings such as resolution, bitrate, and delivery mode. These differences directly affect the overall presentation quality. For every fragment, a Presentation Quality Constraint is specified. This quality represents the lowest acceptable quality level required to maintain instructional clarity. The metadata based on FEMM can be encoded using widely accepted formats like XML or RDF [53]. This ensures flexibility and compatibility with existing systems.

A proof-of-concept prototype, called MoBELearn [41] , was previously developed to demonstrate how CAPS-EMA could be operationalized. It included an adaptation module, a media store with pre-encoded versions, and an RDF-based metadata repository. The present study does not evaluate the MoBELearn system. Instead, it examines the cognitive process involved in the fragment-level metadata annotation required for the CAPS-EMA approach. Specifically, the study investigates whether human experts understand the process of fragmentation based on minimum quality requirements and how they perceive the process of observing instructional video content for identifying fragments and assigning minimum acceptable quality constraints.

Evaluating the Expert-Driven Metadata Authoring Process:

In this section, we discuss the empirical study of the expert-led manual annotation of segment-level metadata in the context of the CAPS-EMA framework. The study was conducted at the School of Electronics and Computer Science, University of Southampton, UK. The specific focus was on the cognitive strategies and perceptions of the experts in

identifying fragment boundaries in educational videos and assigning minimum acceptable quality levels. The study includes observing a manual annotation task in which experienced domain teachers and researchers watch an educational video and make judgments about which portions of the video require the same minimum quality levels in order to preserve critical visual learning content. Experts must assess, for instance, whether reducing quality level would compromise the educational value of a segment or not.

Such tasks of manual annotation of multimedia by experts are not uncommon in the learning technologies domain. From tagging learning objects in MOOCs to aligning transcripts with videos and indexing lecture segments in LMSs, many systems rely on domain experts' insights. This leads to the central research question: Can human experts, based solely on video observation, effectively and cognitively manage the task of segmenting educational videos into instructional fragments and assigning the minimum acceptable presentation quality to each? To address this, we conducted an exploratory observational study where experts independently reviewed and annotated fragments of an educational video. The task of the experts was to identify the lowest quality level at which each segment's instructional value is not compromised. By doing this, we aim to understand the strategies and decision-making patterns used in the manual metadata generation process.

Table 1. Summary of the expert-led metadata authoring study design.

Study element	Description
Study type	Exploratory expert-led observational think-aloud study
Participants	Five experts: three university lecturers and two postdoctoral researchers
Study material	One educational video in four quality versions
Presentation versions	High quality, medium quality, low quality, and audio-only
Authoring task	Identify fragment boundaries and assign minimum acceptable quality for each fragment
Data collected	Think-aloud verbalizations, fragment outputs, MAQ decisions, and post-task questionnaire responses
Analysis	Descriptive thematic coding of decision cues and descriptive summary of questionnaire ratings

Study Objectives:

The goal of this study is to understand the thought process through which experts analyze educational video content and make decisions about fragment boundaries and assigning the lowest acceptable visual quality. The study was conducted without any automated tool. This study simulates a manual annotation process that experts would perform when preparing a CAPS-EMA-compatible resource. These tasks would include segmenting the video into coherent fragments, evaluating the instructional visual content within each fragment, and identifying the lowest presentation quality that would retain the educational clarity for learning.

The study explores four aspects of the mental process involved in metadata annotation: (1) whether experts can successfully identify instructional fragments and assign appropriate quality levels; (2) their understanding of the basis of fragmentation (the visual and instructional cues used to define segment boundaries); (3) their ability to identify the lowest acceptable quality constraint for each fragment; and (4) their perception of the impact of quality reduction on instructional clarity.

This study, therefore, allows us to understand the reasoning processes involved in the manual annotation of segment-level quality constraints. The findings contribute to understanding the human factors that must be considered when designing an authoring system

for metadata generation. The University of Southampton Ethics Committee under reference number 16151 approved the study.

Study Design:

This was an exploratory, expert-led observational study of a manual metadata authoring task. The goal was to examine feasibility and reasoning strategies used by experts while generating the metadata. This study was based on each expert's decisions about fragment boundaries and their assigned minimum acceptable quality (MAQ). We also considered the thought processes they expressed while performing the task. The focus of the analysis was on how decisions were made and the level of cognitive effort involved, rather than evaluating system performance. Table 1 shows the major characteristics of the study design.

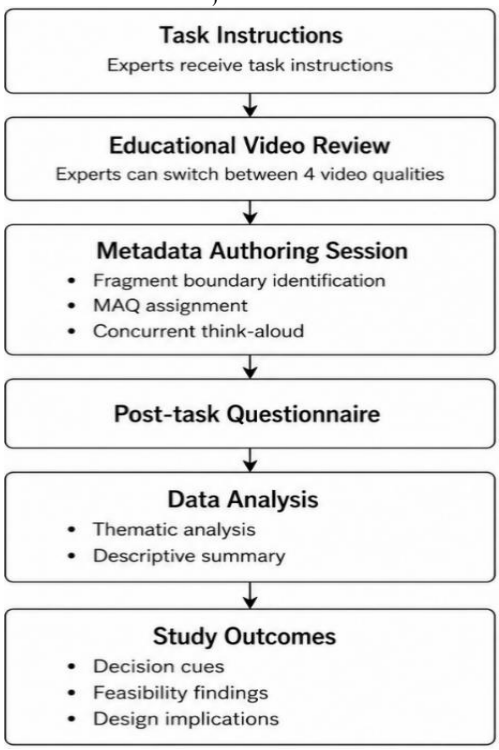

Figure 3. Study Procedure

Participants: The participant group consisted of five academics and researchers with expertise in learning technologies, affiliated with the Electronics and Computer Science Department at the University of Southampton. Three were experienced university lecturers with over five years of teaching experience, and two were postdoctoral researchers specializing in educational tools. Participants were selected on the basis of relevant academic or research experience in learning technologies, university teaching, or educational tools. Given the exploratory nature of the study and common practice in expert/think-aloud studies, we used five experts to capture a range of reasoning patterns rather than to estimate population parameters. Expert evaluations in usability research can be conducted with as few as 3–5 participants [54].

Materials and task: The stimulus was a 13-minute and 15-second slide-based instructional video on the topic of the Semantic Web. The video was previously used in another study within the same project. Participants were provided four representations of the same content: High Quality (124 kbps), Medium Quality (70 kbps), Low Quality (35 kbps), and Audio Only (32 kbps). A simple web interface allowed participants to pause, replay, and switch between versions for visual comparison. The authoring task was to (1) mark start and end times of fragments such that each fragment had a single MAQ requirement, and (2) assign the lowest

quality level that still preserved the perceived visibility and comprehensibility of the fragment's visual instructional information (e.g., text, diagrams, labels).

Procedure: Each session was conducted individually. Participants were briefed on the CAPS-EMA concept, with emphasis on fragment-level constraints and the meaning of MAQ. Participants then watched the video and created a new fragment whenever they judged that the minimum acceptable quality requirement changed (for example, when text/diagram details became important). Participants were encouraged to think aloud throughout. The facilitator used minimal, neutral prompts only when needed for clarification (e.g., "What makes this quality acceptable here?"), without steering decisions.

Data capture: Sessions were recorded for analysis. We retained (i) participants' verbalizations during the task and (ii) the fragment outputs produced by each participant (fragment time ranges and chosen MAQ labels). These data were used to characterize common cues for boundary placement, cues for quality threshold selection, and where variability appeared across experts.

Analysis approach: The think-aloud data were analyzed using descriptive thematic coding focused on authoring decisions. First, we extracted statements in which experts explained why they created a fragment boundary or selected a minimum acceptable quality level. Second, these statements were grouped according to recurring decision cues, such as text legibility, diagram clarity, visual density, narration-only content, layout changes, and borderline quality differences. Third, the grouped cues were reviewed and refined into themes that describe expert reasoning during metadata authoring. The qualitative coding was conducted by one researcher. Therefore, inter-coder reliability was not calculated. The resulting categories were refined iteratively by re-checking the grouped decision statements against the recorded think-aloud data and representative transcript excerpts. The questionnaire responses were summarized descriptively using item-level means because the study was exploratory and involved a small expert sample.

Post-task questionnaire: After completing the authoring task, participants completed a short retrospective questionnaire with seven Likert-scale items (1 = Strongly Disagree to 5 = Strongly Agree) assessing task comprehension, confidence in assigning MAQ, perceived mental effort, difficulty of judging visual sufficiency, perceived simplicity of the overall task, feasibility of applying the process to other videos, and willingness to repeat the task.

Ethics: Participants provided informed consent for participation and recording. The study was approved by the University of Southampton Ethics Committee (Reference: 16151).

This study does not measure energy or battery savings and does not measure learning outcomes. It reports expert judgments and reasoning during manual metadata authoring.

The study procedure is shown in Figure 3.

Findings:

Understanding and Execution of the Annotation Task:

The participants showed a good understanding of the idea and process of fragmentation in the CAPS-EMA framework. They were able to identify fragments, where each fragment had the same minimum quality requirement. For example, Participant 1 stated, "For this part quality 3 is not appropriate but (quality) 2 is better," which shows both understanding and active evaluation. Participant 2 made a similar comment: "Q3 is not bad but Q2 is needed for this information,". Participant 3 pointed out the need for a higher quality level in some cases: "This one should be Quality 1 in high quality". Participant 4 also emphasized this, stating, "Here, the diagram is critical, it's the core of the explanation, so Quality 1 is non-negotiable. Lower resolutions would obscure the labels and defeat the instructional purpose." In the same way, Participant 5 observed, "I noticed that text-heavy slides lose clarity below Quality 2".

All experts reported that the task was manageable. Participant 1 said, "Very easy really easy," while Participant 2 mentioned that difficulty only appeared in borderline cases where quality differences were small. Participant 3 noted, "For this video it was very easy. But for any other video it may be different", suggesting that task difficulty may depend on the type of content. Overall, participants were able to form a mental model of the task and identify the relationship between quality and visual information.

Strategies for Fragment Identification:

We examined how participants decided where to split the video into fragments. Their strategies were mainly based on changes in the content that affected visual quality needs. Participants identified fragments based on: (1) changes in layout such as slide transitions; (2) introduction of new concepts; (3) changes in graphical elements; and (4) changes in screen composition.

Participant 1 focused on text clarity: "Especially the text on labels on the graphs or images… In low quality video there were difficulties." Participant 2 mentioned "change of screen, change of plot" as indicators for creating a new fragment. Participant 5 explained, "I split fragments whenever the RDF graphs appeared, their edge labels and node relationships demanded a dedicated segment. But for slides with just URIs in plain text, I grouped them into a single fragment since the visual demands were uniform."

These responses indicate that participants used a contextual approach combining both visual perception and instructional value when making decisions. Overall, fragment boundaries were not identified simply based on scene changes. Participants also considered impact on the instructional function of the visual content. This suggests that fragment boundary identification is partly pedagogical and semantic, not just a video segmentation task based on scene changes.

Determining Minimum Quality Thresholds:

An important component of metadata generation is identifying a minimum acceptable quality constraint for every fragment. Deciding about the minimum acceptable quality (MAQ) for each fragment requires experts to judge whether the instructional meaning of a fragment is still clear at each reduced quality.

Participants were provided access to four quality versions of the same video through a webpage, including an audio-only version. They could play, pause, and switch between versions for comparison. Their decisions were mainly influenced by the importance of visuals, the amount of graphical detail, and text clarity. Participant 1 explained their approach as "Anything with pictures tends to level two… diagrams or new concepts [require] level one… anything that is descriptive narrative storytelling should be audio only."

Experts showed awareness of how reducing quality may affect visual information. They evaluated text readability, diagram clarity, and overall instructional value by comparing different versions. This process was guided by visual cues and the nature of the visual content.

Participant 1 said, "You can see some blurring on the text there, but that's fine Let's compare it to the above one (higher quality)," showing real-time comparison. For a complex diagram, the same participant said, "This is a diagram and it should be (quality) number 1." Participant 2 also emphasized visual clarity: "Quality 1 for the graph is better," while Participant 3 rejected audio-only for visual content: "This definitely can't be audio… quality 2 for this diagram… I think." Participant 5 added, "The RDF graph's edge labels were unreadable below Quality 2. But when the slide showed just a URI example in plain text, Quality 3 sufficed."

These responses indicate that MAQ assignment was based on perceived instructional sufficiency rather than general visual preference. Experts did not simply choose the visually best version; they attempted to identify the lowest version that still preserved the meaning of the fragment.

Identification of Audio-Only Fragments:

Participants were also able to identify parts of the video that could be delivered as audio-only without losing instructional value. These were usually segments with only verbal explanation. Participant 1 said, "For that last slide… you could use audio as there is nothing (visual) in the slide that could help." Participant 2 noted, "This can be audio as there is nothing special in the visuals," and Participant 3 added, "This one will lose some information but it is ok… So audio is good for this part."

This suggests that audio-only delivery may be acceptable for selected fragments, but only when experts judge that the visual channel does not carry essential instructional information.

Table 2. Minimum acceptable quality selected by each expert for each fragment

Fragment	P1	P2	P3	P4	P5
F1	Q4	Q4	Q4	Q4	Q4
F2	Q1	Q2	Q2	Q2	Q2
F3	Q4	Q4	Q4	Q4	Q4
F4	Q3	Q4	Q3	Q3	Q3
F5	Q4	Q4	Q4	Q4	Q4
F6	Q3	Q2	Q3	Q3	Q2
F7	Q3	Q4	Q2	Q3	Q3
F8	Q4	Q4	Q2	Q4	Q2
F9	Q3	Q2	Q2	Q2	Q3
F10	Q2	Q2	Q2	Q2	Q2
F11	Q1	Q1	Q1	Q1	Q1
F12	Q3	Q2	Q2	Q2	Q2

Variation in Minimum Acceptable Quality Decisions:

The experts did not always select the same minimum acceptable quality for the same fragments. This variation provides additional evidence that MAQ assignment involves expert judgment. The differences were mainly related to how experts judged the importance of visual information and whether instructional meaning remained sufficiently clear at a reduced quality.

Table 2 shows both agreement and variation in MAQ decisions across the five experts. Complete agreement was observed for Fragments 1, 3, 5, 10, and 11, while the remaining fragments showed different levels of variation. Across all 60 decisions, Q2 was selected most frequently (35.0%), followed by Q4 Audio-only (33.3%), Q3 (21.7%), and Q1 (10.0%). This pattern supports the qualitative finding that some fragments produced relatively consistent quality judgments, while others required greater expert interpretation. The recurring decision cues identified during the think-aloud analysis are summarized in Table 3.

Table 3. Expert decision cues observed during segment-level quality metadata authoring.

Theme	Expert cue	Authoring implication
Text legibility	Small text, labels, code, URI examples	Higher quality may be required when text becomes difficult to read
Diagram complexity	Graphs, node-link diagrams, labelled relationships	Higher quality may be required to preserve instructional meaning
Visual density	Many visual elements or detailed slide content	Fragment may need separate treatment and higher quality
Narration dominance	Meaning carried mainly by spoken explanation	Lower quality or audio-only may be acceptable
Layout transition	New slide, new graph, or major visual change	New fragment boundary may be needed

Borderline quality difference	Small difference between adjacent quality levels	More expert judgment and comparison are required

Questionnaire Feedback:

To support the qualitative findings, participants completed a short questionnaire after the task. The aim was to measure their understanding of the task and the cognitive effort involved. The questionnaire included seven items on a 5-point Likert scale (1 = Strongly Disagree, 5 = Strongly Agree). Table 4 reports the item-level mean, standard deviation, and observed response range across the five experts.

Table 4. Post-task expert ratings of the metadata authoring clarity, effort and feasibility

Item	Statement	Mean	SD	Range
S1	The idea of dividing the video into fragments was clear to me.	**4.6**	**0.55**	**4–5**
S2	I felt confident identifying the minimum acceptable quality for each fragment.	**4.2**	**0.45**	**4–5**
S3	Identifying fragments and quality required significant mental effort.	**2.8**	**0.84**	**2–4**
S4	It was mentally challenging to decide if the visual quality was good enough.	**2.4**	**0.55**	**2–3**
S5	I found the overall process to be simple and manageable.	**4.6**	**0.55**	**4–5**
S6	This process would be feasible to apply to other educational videos.	**4.6**	**0.55**	**4–5**
S7	I would be comfortable performing this task for another video.	**4.6**	**0.55**	**4–5**

Descriptively, the highest mean ratings were observed for task clarity, manageability, feasibility, and willingness to repeat the task (S1 and S5–S7; M = 4.6). Confidence in MAQ assignment was also high (S2; M = 4.2). The lower means for the negatively framed effort and difficulty items (S3, M = 2.8; S4, M = 2.4) indicate that the task was not perceived as highly demanding by this expert group. Given the sample of five experts, these values are interpreted descriptively rather than as population-level evidence.

High agreement was also observed for task simplicity and feasibility. Items S5, S6, and S7 all received mean scores of 4.6. This shows that participants found the task manageable and were comfortable performing it again. Overall, these ratings provide preliminary evidence that the authoring task was perceived as manageable by this expert group, while still requiring some cognitive effort.

Discussion:

The findings suggest that expert-led segment-level quality metadata authoring is feasible for the studied video and expert group. Experts were able to understand the authoring task, identify instructional fragments, and assign minimum acceptable quality levels. The post-task ratings also indicated that the process was perceived as clear and manageable. Accordingly, the findings should be interpreted as exploratory evidence from five experts working with one educational video and should not be generalized to other expert populations or video types.

Metadata Authoring as an Interpretive Task:

The study shows that metadata authoring for learning-sensitive adaptation is not a purely mechanical segmentation task. Experts did not only look for scene transitions or time intervals. They also considered the instructional role of the visual content. For example, fragments containing labels, diagrams, graphs, or small text were treated differently from fragments dominated by narration or simple slides. This suggests that fragment-level quality metadata requires both visual inspection and instructional judgment.

The variation in fragment boundaries and quality thresholds also supports this interpretation. Experts generally followed similar reasoning patterns, but their exact boundaries and quality choices were not always identical. This is expected because perceived visual comprehensibility can depend on the expert's interpretation of the content, the importance assigned to visual elements, and the perceived redundancy between visual and spoken information.

Cues Used in Expert Decision-Making:

The think-aloud data indicate that experts used several recurring cues when making authoring decisions. Text legibility was one of the strongest cues. When labels, code, URI examples, or slide text became difficult to read, experts tended to assign higher quality levels. Diagram complexity was another important cue. Graphs, node-link diagrams, and labelled relationships required higher quality because loss of detail could affect understanding.

Experts also considered visual redundancy:

When the visual channel did not add much beyond narration, lower quality or audio-only delivery was considered acceptable. This finding is important because it suggests that learning-sensitive adaptation should not only detect visual complexity. It should also consider whether the visual content is instructionally necessary.

Implications for Semi-Automated Authoring Tools:

The findings suggest that future metadata authoring tools should support expert judgment rather than simply replace it. Semi-automated tools could assist by detecting slide changes, scene boundaries, text-heavy segments, diagram regions, and visual density. They could also provide side-by-side quality comparison and allow experts to quickly mark the lowest acceptable quality.

A useful authoring tool could therefore include several functions: automatic candidate segmentation, visual cue detection, preview of multiple quality versions, audio-only eligibility marking, and confidence indicators for difficult cases. Such tools could be evaluated for their potential to reduce manual effort while retaining expert oversight of instructional judgments.

Limits of Full Automation:

The study also highlights why fully automated metadata generation remains challenging. Educational importance is not always the same as visual complexity. A visually simple slide may contain a key definition, while a visually complex slide may contain decorative or redundant information. Automated systems may detect complexity, saliency, or text density, but they may not reliably determine whether a visual element is instructionally necessary.

Therefore, fully automated metadata generation should be treated as a future research direction. Although advances in artificial intelligence may assist with tasks such as scene detection, text recognition, diagram identification, and candidate fragment generation, instructional importance cannot always be inferred reliably from visual characteristics alone. In the present authoring task, human judgment was used to determine whether visual information was considered important for perceived instructional understanding. A more practical near-term approach is to develop semi-automated human-in-the-loop authoring environments that assist experts by proposing fragment boundaries, highlighting visually demanding content, and supporting quality comparison while leaving the final instructional decisions under expert control. These findings therefore motivate future evaluation of semi-automated human-in-the-loop authoring as a possible approach to learning-sensitive metadata generation.

Limitations and Future Work:

This study has several limitations. First, it involved only five experts. This sample size is acceptable for an exploratory expert-led think-aloud study, but it does not support statistical generalization. The goal was to identify reasoning patterns and feasibility signals rather than estimate population-level effects.

Second, all experts were drawn from a similar academic and research environment. Future work should include more diverse annotator groups.

Third, the study used one educational video. The authoring task may be more difficult for other types of educational multimedia, such as coding demonstrations, laboratory demonstrations, classroom recordings, animations, or fast-paced procedural videos. The findings should therefore be interpreted as evidence from one controlled authoring context.
Fourth, the study did not quantify agreement between experts. Although some variation in fragment boundaries and quality choices was observed, future work may examine agreement more systematically.

Fifth, the study did not compare manual authoring with semi-automated or automated metadata generation. Future studies should compare expert-only authoring, tool-assisted authoring, and machine-generated suggestions.

Future work should therefore examine semi-automated authoring tools, agreement between experts, scalability across different video types, and the effect of expert-generated metadata on learner experience and energy-aware multimedia delivery.

Conclusion:

This paper presented an exploratory expert-led study of manual segment-level quality metadata authoring for learning-sensitive educational video adaptation. The study focused on the human reasoning required to identify instructional fragments and assign minimum acceptable quality levels.

The findings suggest that experts were able to understand and perform the metadata authoring task for the sample video. They used practical visual and instructional cues, including text legibility, diagram complexity, visual density, narration-only content, and layout changes. These cues helped them decide where fragments should begin and end and what minimum quality level was needed to preserve perceived visual comprehensibility.

Post-task ratings were consistent with the think-aloud observations, indicating high perceived task clarity and manageability while the variation in fragment boundaries and MAQ choices showed that the authoring process remained interpretive.

The main contribution of this paper is not a validated adaptation system or an energy-saving result. Rather, the contribution is human-factors evidence about the expert reasoning involved in fragment-level quality metadata authoring. These findings can inform the design of future semi-automated authoring tools that support expert judgment through candidate segmentation, visual cue detection, quality preview, and confidence marking.

Future research should evaluate this authoring process across more diverse video types, larger expert groups, and semi-automated tool-supported workflows. Further work is also needed to examine the impact of the expert-generated metadata decisions on learner experience in implemented systems.

References:

[1] M. Sharples, J. Taylor, and G. Vavoula, "A Theory of Learning for the Mobile Age," *Medien. neuen Kult.*, pp. 87–99, 2010, doi: 10.1007/978-3-531-92133-4_6.

[2] S. K. S. Cheung, "A study on the use of mobile devices for distance learning," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 7411 LNCS, pp. 89–98, 2012, doi: 10.1007/978-3-642-32018-7_9/Save-Research.

[3] A.-N. Moldovan and C. H. Muntean, "Towards Personalised and Adaptive Multimedia in M-learning Systems," *Proc. E-Learn 2011--World Conf. E-Learning Corp. Gov. Heal. High. Educ.*, pp. 782–791, 2011, Accessed: Aug. 26, 2026. [Online]. Available: https://www.learntechlib.org/primary/p/38807/

[4] E. Navarrete, A. Nehring, S. Schanze, R. Ewerth, and A. Hoppe, "A Closer Look into Recent Video-based Learning Research: A Comprehensive Review of Video Characteristics, Tools, Technologies, and Learning Effectiveness," *Int. J. Artif. Intell.*

Educ. 2025 354, vol. 35, no. 4, pp. 1631–1694, May 2025, doi: 10.1007/S40593-025-00481-X.

[5] C. Zhang, Z. Wang, Z. Fang, and X. Xiao, "Guiding student learning in video lectures: Effects of instructors' emotional expressions and visual cues," *Comput. Educ.*, vol. 218, p. 105062, Sep. 2024, doi: 10.1016/J.COMPEDU.2024.105062.

[6] M. A. Hoque, M. Siekkinen, and J. K. Nurminen, "Energy efficient multimedia streaming to mobile devices-A survey," *IEEE Commun. Surv. Tutorials*, vol. 16, no. 1, pp. 579–597, Mar. 2014, doi: 10.1109/SURV.2012.111412.00051.

[7] R. Trestian, A. N. Moldovan, O. Ormond, and G. M. Muntean, "Energy consumption analysis of video streaming to Android mobile devices," *Proc. 2012 IEEE Netw. Oper. Manag. Symp. NOMS 2012*, pp. 444–452, 2012, doi: 10.1109/NOMS.2012.6211929.

[8] Q. Tan, X. Zhang, and R. McGreal, "The 5 R Adaptation Framework for Location-Based Mobile Learning Systems," 2011.

[9] A. N. Moldovan, S. Weibelzahl, and C. H. Muntean, "Energy-aware mobile learning:Opportunities and challenges," *IEEE Commun. Surv. Tutorials*, vol. 16, no. 1, pp. 234–265, Mar. 2014, doi: 10.1109/SURV.2013.071913.00194.

[10] J. Zhang, D. Wu, S. Ci, H. Wang, and A. K. Katsaggelos, "Power-aware mobile multimedia: A survey (Invited Paper)," *J. Commun.*, vol. 4, no. 9, pp. 600–613, 2009, doi: 10.4304/JCM.4.9.600-613.

[11] M. A. Hoque, M. Siekkinen, J. K. Nurminen, M. Aalto, and S. Tarkoma, "Mobile multimedia streaming techniques: QoE and energy saving perspective," *Pervasive Mob. Comput.*, vol. 16, no. PA, pp. 96–114, Jan. 2015, doi: 10.1016/J.PMCJ.2014.05.004.

[12] C. J. Lian, S. Y. Chien, C. P. Lin, P. C. Tseng, and L. G. Chen, "Power-aware multimedia: Concepts and design perspectives," *IEEE Circuits Syst. Mag.*, vol. 7, no. 2, pp. 26–34, 2007, doi: 10.1109/MCAS.2007.4299440.

[13] A. N. Moldovan and C. H. Muntean, "Subjective assessment of bitdetecta mechanism for energy-aware multimedia content adaptation," *IEEE Trans. Broadcast.*, vol. 58, no. 3, pp. 480–492, 2012, doi: 10.1109/TBC.2012.2191688.

[14] D. Kelly, J. R. Delannoy, J. McDonald, C. Markham, and J. M. Donald, "Assessing Power Saving Techniques and Their Impact on E-learning Users," *China Irel. Int. Conf. Inf. Commun. Technol.*, pp. 181–187, 2009.

[15] "Energy-Aware Streaming Multimedia Adaptation: An Educational Perspective | Request PDF." Accessed: Aug. 26, 2026. [Online]. Available: https://www.researchgate.net/publication/299448882_Energy-Aware_Streaming_Multimedia_Adaptation_An_Educational_Perspective

[16] X. Ma *et al.*, "QAVA: QoE-Aware Adaptive Video Bitrate Aggregation for HTTP Live Streaming Based on Smart Edge Computing," *IEEE Trans. Broadcast.*, vol. 68, no. 3, pp. 661–676, Sep. 2022, doi: 10.1109/TBC.2022.3171131.

[17] M. H. Bin Waheed *et al.*, "A New Efficient Architecture for Adaptive Bit-Rate Video Streaming," *Sustain. 2021, Vol. 13, Page 4541*, vol. 13, no. 8, p. 4541, Apr. 2021, doi: 10.3390/SU13084541.

[18] E. İren and A. Kantarci, "Content Aware Video Streaming with MPEG DASH Technology," *TEM J.*, vol. 11, no. 2, pp. 611–619, May 2022, doi: 10.18421/TEM112-15.

[19] W. ur Rahman and E. N. Huh, "Content-aware QoE optimization in MEC-assisted Mobile video streaming," *Multimed. Tools Appl. 2023 8227*, vol. 82, no. 27, pp. 42053–42085, Apr. 2023, doi: 10.1007/S11042-023-15163-W.

[20] A. Zabrovskiy, P. Agrawal, C. Timmerer, and R. Prodan, "FAUST: Fast Per-Scene Encoding Using Entropy-Based Scene Detection and Machine Learning," *Conf. Open*

Innov. Assoc. Fruct, vol. 2021-October, pp. 267–278, 2021, doi: 10.23919/FRUCT53335.2021.9599963.

[21] A. Neparidze *et al.*, "Deep Explainable Content-Aware Per-Scene Video Encoding," *2024 Int. Conf. Comput. Netw. Commun. ICNC 2024*, pp. 490–494, 2024, doi: 10.1109/ICNC59896.2024.10555944.

[22] G. Gao, L. Dong, H. Zhang, Y. Wen, and W. Zeng, "Content-Aware Personalised Rate Adaptation for Adaptive Streaming via Deep Video Analysis," *IEEE Int. Conf. Commun.*, vol. 2019-May, May 2019, doi: 10.1109/ICC.2019.8761156.

[23] J. A. Ghauri, S. Hakimov, and R. Ewerth, "Classification of Important Segments in Educational Videos using Multimodal Features," *CEUR Workshop Proc.*, vol. 2699, Oct. 2020, Accessed: Aug. 26, 2026. [Online]. Available: https://arxiv.org/pdf/2010.13626

[24] S. A. Jalal, N. Gibbins, D. Millard, B. Al-Hashimi, and N. R. Aljohani, "Content-aware power saving multimedia adaptation for mobile learning," *Int. Conf. Next Gener. Mob. Appl. Serv. Technol.*, pp. 256–261, Nov. 2013, doi: 10.1109/NGMAST.2013.53.

[25] C. H. Lin, J. C. Liu, and C. W. Liao, "Energy analysis of multimedia video decoding on mobile handheld devices," *Proc. - 2007 Int. Conf. Multimed. Ubiquitous Eng. MUE 2007*, pp. 120–125, 2007, doi: 10.1109/MUE.2007.125.

[26] T. Stockhammer, "Dynamic adaptive streaming over HTTP - Standards and design principles," *MMSys'11 - Proc. 2011 ACM Multimed. Syst. Conf.*, pp. 133–143, Feb. 2011, doi: 10.1145/1943552.1943572.

[27] K. Spiteri, R. Sitaraman, and D. Sparacio, "From theory to practice: Improving Bitrate adaptation in the DASH reference player," *ACM Trans. Multimed. Comput. Commun. Appl.*, vol. 15, no. 2s, Aug. 2019, doi: 10.1145/3336497;TAXONOMY:TAXONOMY:ACM-PUBTYPE;PAGEGROUP:STRING:PUBLICATION.

[28] O. Machidon, J. Asprov, T. Fajfar, and V. Pejović, "Context-aware adaptation of mobile video decoding resolution," *Multimed. Tools Appl. 2022 8212*, vol. 82, no. 12, pp. 17599–17630, Oct. 2022, doi: 10.1007/S11042-022-13787-Y.

[29] J. Meng, Q. Xu, and Y. C. Hu, *Proactive {Energy-Aware} Adaptive Video Streaming on Mobile Devices*. 2021. Accessed: Aug. 26, 2026. [Online]. Available: https://www.usenix.org/conference/atc21/presentation/meng

[30] C. Yue, S. Sen, B. Wang, Y. Qin, and F. Qian, "Energy considerations for ABR video streaming to smartphones: Measurements, models and insights," *MMSys 2020 - Proc. 2020 Multimed. Syst. Conf.*, pp. 153–165, May 2020, doi: 10.1145/3339825.3391867;CTYPE:STRING:BOOK.

[31] C. Diaz, A. Fernandez, F. Sacristan, and N. Garcia, "Energy-and Quality-Aware Video Request Policy for Wireless Adaptive Streaming Clients," *IEEE Trans. Consum. Electron.*, vol. 66, no. 4, pp. 366–375, Nov. 2020, doi: 10.1109/TCE.2020.3034619.

[32] G. Kim, D. Lee, and M. Song, "Design and Implementation of Bitrate Adaptation Schemes for Power Capping in Wi-Fi Video Streaming," *IEEE Access*, vol. 8, pp. 22581–22591, 2020, doi: 10.1109/ACCESS.2020.2969696.

[33] C. G. Bampis, Z. Li, A. K. Moorthy, I. Katsavounidis, A. Aaron, and A. C. Bovik, "Study of temporal effects on subjective video quality of experience," *IEEE Trans. Image Process.*, vol. 26, no. 11, pp. 5217–5231, Nov. 2017, doi: 10.1109/TIP.2017.2729891.

[34] B. O. Turkkan *et al.*, "GreenABR: Energy-Aware Adaptive Bitrate Streaming with Deep Reinforcement Learning," *MMSys 2022 - Proc. 13th ACM Multimed. Syst. Conf.*, pp. 150–163, Aug. 2022, doi: 10.1145/3524273.3528188;WGROUP:STRING:ACM.

[35] A. Bunt, G. Carenini, and C. Conati, "Adaptive content presentation for the Web,"

Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 4321 LNCS, pp. 409–432, 2007, doi: 10.1007/978-3-540-72079-9_13/SAVE-RESEARCH.

[36] M. Venkata Phani Kumar, C. Timmerer, and H. Hellwagner, "Mipso: Multi-period per-scene optimization for http adaptive streaming," *Proc. - IEEE Int. Conf. Multimed. Expo*, vol. 2020-July, Jul. 2020, doi: 10.1109/ICME46284.2020.9102775.

[37] M. Bhat, J. M. Thiesse, and P. Le Callet, "A case study of machine learning classifiers for real-time adaptive resolution prediction in video coding," *Proc. - IEEE Int. Conf. Multimed. Expo*, vol. 2020-July, Jul. 2020, doi: 10.1109/ICME46284.2020.9102934.

[38] N. Souane, M. Bourenane, and Y. Douga, "Deep Reinforcement Learning-Based Approach for Video Streaming: Dynamic Adaptive Video Streaming over HTTP," *Appl. Sci. 2023, Vol. 13, Page 11697*, vol. 13, no. 21, p. 11697, Oct. 2023, doi: 10.3390/APP132111697.

[39] V. V. Menon, J. Zhu, P. T. Rajendran, H. Amirpour, P. Le Callet, and C. Timmerer, "Just Noticeable Difference-Aware Per-Scene Bitrate-Laddering for Adaptive Video Streaming," *Proc. - IEEE Int. Conf. Multimed. Expo*, vol. 2023-July, pp. 1673–1678, 2023, doi: 10.1109/ICME55011.2023.00288.

[40] Y. Lu *et al.*, "ABUV: Adaptive bitrate and upsampling for video streaming on mobile devices," *Comput. Networks*, vol. 257, p. 110994, Feb. 2025, doi: 10.1016/J.COMNET.2024.110994.

[41] S. A. Jalal, N. Gibbins, D. Millard, B. Al-Hashimi, and N. R. Aljohani, "Energy-aware adaptation of educational multimedia in mobile learning," *ACM Int. Conf. Proceeding Ser.*, pp. 434–439, 2013, doi: 10.1145/2536853.2536896;PAGE:STRING:ARTICLE/CHAPTER.

[42] A. N. Moldovan, C. A. Moraru, and C. H. Muntean, "MediaMTool: Multimedia content management tool," *IEEE Int. Symp. Broadband Multimed. Syst. Broadcast. BMSB 2011 - Conf. Program.*, 2011, doi: 10.1109/BMSB.2011.5954907.

[43] A. N. Moldovan and C. H. Muntean, "DQAMLearn: Device and QoE-Aware Adaptive Multimedia Mobile Learning Framework," *IEEE Trans. Broadcast.*, vol. 67, no. 1, pp. 185–200, Mar. 2021, doi: 10.1109/TBC.2020.3028338.

[44] A. Molnar, "Content type and perceived multimedia quality in mobile learning," *Multimed. Tools Appl. 2016 7620*, vol. 76, no. 20, pp. 21613–21627, Nov. 2016, doi: 10.1007/S11042-016-4062-2.

[45] A. Molnar, "The effect of content-type and video adaptation on information assimilation," *ACM Int. Conf. Proceeding Ser.*, pp. 123–127, Nov. 2018, doi: 10.1145/3282353.3282354;PAGEGROUP:STRING:PUBLICATION.

[46] M. McClelland, "Metadata Standards for Educational Resources," *Computer (Long. Beach. Calif).*, vol. 36, no. 11, pp. 107–109, Nov. 2003, doi: 10.1109/MC.2003.1244540.

[47] P. Ingavélez-Guerra, S. Otón-Tortosa, J. Hilera-González, and M. Sánchez-Gordón, "The use of accessibility metadata in e-learning environments: a systematic literature review," *Univers. Access Inf. Soc. 2021 222*, vol. 22, no. 2, pp. 445–461, Dec. 2021, doi: 10.1007/S10209-021-00851-X.

[48] W. Simao De Deus and E. Francine Barbosa, "The Use of Metadata in Open Educational Resources Repositories: An Exploratory Study," *Proc. - 2020 IEEE 44th Annu. Comput. Software, Appl. Conf. COMPSAC 2020*, pp. 123–132, Jul. 2020, doi: 10.1109/COMPSAC48688.2020.00025.

[49] P. S. Chiu, H. C. Chen, Y. M. Huang, C. J. Liu, M. C. Liu, and M. H. Shen, "A video annotation learning approach to improve the effects of video learning," *Innov. Educ. Teach. Int.*, vol. 55, no. 4, pp. 459–469, Jul. 2018, doi:

10.1080/14703297.2016.1213653.
[50] "(PDF) Characterising Video Segments to Support Learning." Accessed: Aug. 26, 2026. [Online]. Available: https://www.researchgate.net/publication/346975350_Characterising_Video_Segments_to_Support_Learning
[51] B. Memarian and T. Doleck, "Human-in-the-loop in artificial intelligence in education: A review and entity-relationship (ER) analysis," *Comput. Hum. Behav. Artif. Humans*, vol. 2, no. 1, p. 100053, Jan. 2024, doi: 10.1016/J.CHBAH.2024.100053.
[52] C. N. C. Lam and H. Habil, "The Use of Video Annotation in Education: A Review," *Asian J. Univ. Educ.*, vol. 17, no. 4, pp. 84–94, Oct. 2021, doi: 10.24191/AJUE.V17I4.16208.
[53] E. Duval, W. Hodgins, S. Sutton, and S. L. Weibel, "Metadata Principles and Practicalities," *D-Lib Mag.*, vol. 8, no. 4, Apr. 2002, doi: 10.1045/APRIL2002-WEIBEL.
[54] "Usability Engineering - Jakob Nielsen - Google Books." Accessed: Aug. 27, 2026. [Online]. Available: https://books.google.com.pk/books?id=95As2OF67f0C&printsec=copyright&redir_esc=y#v=onepage&q&f=false

Copyright © by authors and 50Sea. This work is licensed under Creative Commons Attribution 4.0 International License.