JOURNAL OF INTERNATIONAL SCIENTIFIC RESEARCH
Volume 4, Issue 9, October-November 2026 Online ISSN: 3030-3508
https://spaceknowladge.com

DIFFERENTIATING COGNITIVE COMPLEXITY IN AUDIOVISUAL TASKS FOR RUSSIAN LANGUAGE TEACHING WITHIN AN EDUCATIONAL CLUSTER: A METHODOLOGICAL FRAMEWORK

Aziz Avazovich Umarov
Dean of the Faculty of Languages,
Associate Professor, PhD, Termiz State Pedagogical Institute,
Termez, Uzbekistan
E-mail: azizumarov38239@gmail.com
ORCID: 0009-0000-7223-8059

ИНФОРМАЦИЯ О СТАТЬЕ

ИСТОРИЯ СТАТЬИ:

Received:30.10.2026
Revised: 31.10.2026
Accepted:01.11.2026

КЛЮЧЕВЫЕ СЛОВА:

Russian language teaching; audiovisual tasks; cognitive complexity; educational cluster; cognitive load; multimedia learning; task sequencing; differentiation.

АННОТАЦИЯ:

This study develops a methodological framework for differentiating the cognitive complexity of audiovisual tasks in Russian language teaching within an educational cluster. The problem addressed is not the general effectiveness of audiovisual instruction, but the absence of a transparent procedure for matching the cognitive demands of video- and audio-based tasks to learners' language proficiency, prior knowledge, and processing capacity. The study adopts an integrative methodological design combining recent research on multimedia learning, cognitive load, audiovisual second-language input, task complexity, and task sequencing. The resulting framework distinguishes four levels of audiovisual task complexity-recognition, guided integration, inferential reconstruction, and productive transfer-and operationalizes complexity through five variables: linguistic density, number of interacting information elements, degree of visual–verbal correspondence, inferential demand, and response autonomy. The educational cluster functions as an organizational mechanism for shared media banks, common task descriptors, teacher calibration, and cross-institutional quality assurance. The proposed framework shifts attention from merely selecting audiovisual content to engineering a progression of

cognitively calibrated learning actions. It also provides a basis for formative assessment and for future experimental validation in Russian as a second/foreign language contexts. The model is particularly relevant where learner groups are heterogeneous and where audiovisual resources are used across schools, teacher-training institutions, and methodological centers..

INTRODUCTION

Audiovisual instruction has become a stable component of contemporary second- and foreign-language pedagogy because it combines spoken language with facial expression, gesture, setting, movement, written text, and other contextual cues. Recent syntheses show that such input can support comprehension and language learning, but its effect is conditional rather than automatic: captions, visual relevance, learner proficiency, speech rate, and the relation between auditory and visual information may either facilitate processing or increase cognitive demands (Montero Perez, 2022; Tan, Samad, & Ismail, 2022; Muñoz, 2022). Consequently, the central methodological question is no longer whether video should be used, but how audiovisual tasks should be designed so that the amount and interaction of information correspond to learners' available processing resources.

This issue is especially important in Russian language teaching within an educational cluster. A cluster connects schools, higher education institutions, methodological centers, and teacher-development structures around common educational goals. Such an environment makes it possible to share audiovisual resources and methodological expertise, yet it also creates a standardization problem: the same video fragment can be cognitively simple for one learner group and excessively demanding for another. Previous work on audiovisual Russian-language instruction has addressed methodological possibilities, integration with traditional methods, teacher roles, communicative competence, intercultural competence, and principles of content selection. A distinct problem remains underdeveloped: the systematic differentiation of the cognitive complexity of the learning tasks constructed around audiovisual input.

The purpose of this article is therefore to formulate a methodological framework for differentiating audiovisual tasks according to cognitive complexity and to define how the framework can be implemented through educational-cluster mechanisms. Three questions guide the study: (1) which variables make an audiovisual language task cognitively simple or complex; (2) how can task complexity be sequenced without reducing authenticity; and (3) what functions should a cluster perform to ensure consistent design, assessment, and

adaptation of such tasks? The contribution lies in moving from content-centered selection toward task-centered cognitive calibration.

MATERIALS AND METHODS

The study uses an integrative methodological synthesis rather than an experimental design. Its evidential base consists of peer-reviewed research on four intersecting domains: the cognitive theory of multimedia learning, cognitive load and task complexity, audiovisual input in second-language acquisition, and task sequencing in language pedagogy. The framework draws particularly on the dual-channel, limited-capacity, and active-processing assumptions of multimedia learning (Mayer, 2021, 2024), on the interpretation of task complexity through element interactivity (Chen, Paas, & Sweller, 2023), and on recent evidence concerning subtitles, visual information, listening comprehension, and learner proficiency in audiovisual learning (Montero Perez, 2022; Shin & Lee, 2024).

For methodological purposes, an audiovisual task is defined here as a pedagogically structured action in which a learner must process an audio or video stimulus and produce an observable response. Complexity is distinguished from difficulty. Difficulty is the learner's experienced effort or success rate; complexity refers to the structural and cognitive demands built into the task. This distinction is critical because a task with few elements can still be difficult for a beginner, while a task containing many elements may be manageable for an advanced learner with strong prior knowledge. The framework therefore treats task complexity as a relation between task architecture and learner readiness rather than as a fixed property of the media resource.

Five variables were selected for operationalization. The first is linguistic density: speech rate, lexical frequency, grammatical compression, and discourse length. The second is element interactivity: the number of pieces of information that must be held and coordinated simultaneously. The third is visual–verbal correspondence: whether images directly clarify speech, provide only partial support, or introduce competing information. The fourth is inferential demand: the extent to which meaning must be reconstructed from implication, pragmatics, temporal relations, or cultural context. The fifth is response autonomy: whether learners select a supplied answer, reconstruct information with prompts, or independently formulate an oral or written response.

The framework also incorporates two adaptation principles. First, complexity should be regulated through task design before the audiovisual material itself is simplified. Authentic input can often be retained while prompts, segmentation, captioning, replay conditions, note-taking support, and response formats are differentiated. Second, sequencing should follow a controlled progression from lower to higher integration demands. A systematic review of task complexity and sequencing research indicates that simple-to-complex progression can support L2 performance, although the effect depends on how complexity is operationalized

(Jin & Yan, 2025). The present model therefore avoids a single global "easy–hard" scale and instead uses multiple adjustable dimensions.

CONCEPTUAL BASIS

Multimedia learning theory explains why audiovisual instruction requires careful calibration. Learners process verbal/auditory and visual/pictorial information through partially distinct channels, but each channel has limited capacity. Effective learning requires selecting relevant information, organizing it into coherent mental representations, and integrating it with prior knowledge (Mayer, 2024). In language learning, these operations occur simultaneously with lexical recognition, syntactic parsing, pragmatic interpretation, and monitoring of comprehension. An apparently short video can therefore generate substantial load when speech is fast, the visual scene changes rapidly, captions compete for attention, and the learner must infer unstated relations.

Empirical research supports a conditional view of multimedia support. Subtitles can facilitate comprehension in some contexts, yet video itself does not automatically improve understanding; benefits depend on learner characteristics and attentional distribution. Proficiency can also moderate the extent to which subtitles or visual cues reduce extraneous load. For Russian-language instruction, this means that audiovisual enrichment should not be equated with cognitive accessibility. The teacher must regulate the number of simultaneous demands and determine which cues are supportive at a given learning stage.

RESULTS

The methodological synthesis resulted in a four-level model. The levels describe the dominant cognitive operation required from the learner rather than the formal type of media. The same video fragment can therefore support several levels of work when the task architecture changes.

Complexity level	Dominant cognitive demand	Typical audiovisual task	Cluster-level support
Level I: Recognition	Identify explicit words, facts, speakers, or visible actions; low inferential demand.	Select a matching image; identify a stated fact; mark key lexical items.	Common beginner task templates and calibrated media segments.
Level II: Guided integration	Coordinate auditory and visual information with structured prompts.	Complete a sequence; match utterances to motives; fill a partially structured note grid.	Shared scaffolds, caption variants, and teacher calibration.

Level III: Inferential reconstruction	Integrate dispersed cues, infer relations, interpret pragmatic or cultural meaning.	Explain an implied intention; reconstruct cause–effect relations; compare viewpoints.	Cross-institutional discussion cases and expert-reviewed task banks.
Level IV: Productive transfer	Use processed audiovisual information to solve a new communicative problem independently.	Produce a summary, role-play, argument, mediation task, or alternative scenario.	Common rubrics, moderation, and portfolio-based assessment across cluster partners.

At Level I, complexity is reduced by maximizing direct correspondence between speech and visual context, limiting the number of required information elements, and using constrained responses. The objective is not passive viewing but stable perceptual anchoring. At Level II, learners begin to combine channels: they may connect an utterance with a gesture, arrange events in sequence, or use captions selectively to verify uncertain lexical forms. The task remains scaffolded, but the learner must integrate more than one source of evidence.

Level III requires inferential reconstruction. Learners must identify implicit motives, distinguish literal from pragmatic meaning, connect dispersed details, or interpret culturally conditioned behavior. Here, the teacher should not remove all uncertainty; instead, the task should make the inferential route manageable by controlling segment length, replay policy, pre-viewing activation, and the number of simultaneous questions. At Level IV, the audiovisual material becomes a resource for independent language production. Learners transform information by summarizing, mediating, debating, role-playing, problem-solving, or generating an alternative continuation. The principal complexity factor is response autonomy, because the learner must choose linguistic resources and organize them without a fully predetermined response structure.

Progression between levels should be adaptive rather than strictly linear. A learner may perform Level III inference successfully with strong visual support but require Level II scaffolding when the same inference is based primarily on rapid speech. The framework therefore recommends a “complexity profile” instead of a single score. Teachers can rate each task on the five variables-linguistic density, element interactivity, visual–verbal correspondence, inferential demand, and response autonomy-and adjust one or two dimensions at a time. This preserves diagnostic value: when performance declines, the teacher can identify which dimension exceeded the learner’s current processing capacity.

The educational cluster should be understood as an infrastructure for methodological coordination, not merely as a network for distributing digital materials. Four functions are essential. First, a shared audiovisual media bank should store not only files or links, but metadata describing language level, speech rate, duration, lexical density, cultural content,

caption availability, and potential cognitive load. This converts a media repository into a pedagogically searchable resource.

Second, the cluster should maintain a common task-descriptor system. Each task can be tagged by the five complexity variables and by the four model levels. Such descriptors allow teachers in different institutions to compare tasks meaningfully and to reuse materials without assuming that identical media automatically generate identical learning demands. Third, teacher calibration is required. Short moderation sessions in which teachers independently classify sample tasks and then discuss differences can improve the reliability of complexity judgments. This is particularly important because prior knowledge and learner characteristics influence the experienced load of a task.

Fourth, common formative-assessment rubrics should distinguish comprehension failure from task-overload failure. For example, an incorrect answer after a dense video may reflect insufficient lexical knowledge, poor attentional allocation, excessive inferential demand, or inability to formulate a response despite adequate comprehension. A cluster-wide rubric can therefore include indicators for information extraction, integration of auditory and visual cues, inference, linguistic formulation, and independent transfer. These indicators make it possible to adapt the subsequent task rather than simply assign a lower mark.

The framework also supports differentiated instruction in heterogeneous groups. Instead of preparing completely different videos for every proficiency subgroup, teachers can retain a common audiovisual core and vary scaffolding. Beginners may receive segmented viewing, keyword previews, controlled captions, and recognition tasks; intermediate learners may receive reduced scaffolding and integration questions; advanced learners may work with uncaptioned or partially captioned input and complete inferential or productive-transfer tasks. This approach is organizationally realistic for a cluster because a single validated media object can support several differentiated task packages.

DISCUSSION

The proposed model extends audiovisual methodology in three respects. First, it treats task design as the primary unit of differentiation. This is consistent with evidence that the value of audiovisual input depends on how learners allocate attention and how verbal and visual information interact, rather than on the mere presence of video. Second, the model links L2 task complexity with cognitive load through element interactivity. Chen, Paas, and Sweller (2023) emphasize that complexity depends both on the structure of information and on knowledge stored in long-term memory. Applied to Russian language teaching, the implication is that identical language input can impose different intrinsic loads on learners with different lexical, grammatical, and sociocultural knowledge.

Third, the model assigns the educational cluster a quality-assurance role. Shared resources are pedagogically useful only when institutions also share descriptors, calibration procedures, and assessment criteria. This prevents two common problems: technological inflation, in

which more multimedia is assumed to mean better learning, and unstructured differentiation, in which teachers simplify tasks intuitively without documenting what was changed. A cognitively calibrated task bank offers a more transparent basis for instructional continuity between institutions and for replication in future research.

The framework is intended to be empirically testable. A validation study should separate three forms of evidence: design validity, implementation reliability, and learning outcomes. Design validity can be examined by asking independent experts to rate audiovisual tasks on the five proposed variables and then estimating inter-rater agreement. The aim is not to force complete consensus, but to determine whether the descriptors are sufficiently explicit for different teachers to reach comparable judgments about linguistic density, element interactivity, visual–verbal correspondence, inferential demand, and response autonomy.

Implementation reliability should then be studied in authentic cluster settings. Teachers from different partner institutions can use the same audiovisual core and the same complexity descriptors while adapting scaffolds to their learner groups. Lesson observations and teacher logs can document whether intended task levels are maintained in practice. This step is essential because task complexity can be altered unintentionally: additional explanations may simplify inference, unrestricted replay may reduce memory demands, or an apparently supportive worksheet may introduce competing visual information. A cluster-based model is methodologically meaningful only if its design principles survive classroom adaptation.

Outcome validation should compare learner performance across calibrated task levels and proficiency groups. Suitable indicators include comprehension accuracy, response completeness, lexical and syntactic adequacy, perceived mental effort, delayed retention, and the quality of transfer to a new communicative situation. The model predicts neither that more complex tasks are always superior nor that lower cognitive load is always desirable. Instead, productive learning should occur when task demands are challenging enough to require active integration but remain within a range that permits successful processing. This expectation is consistent with contemporary cognitive-load research, which distinguishes useful effort from overload and emphasizes the interaction between task structure and prior knowledge.

A quasi-experimental cluster design could therefore use equivalent learner groups completing the same audiovisual material under differently calibrated task conditions. Such research should report the exact task manipulations rather than only naming a method as ‘audiovisual’. For example, investigators should specify segment duration, caption status, number of required information elements, inferential operations, replay conditions, and response format. Transparent reporting would make results replicable and would permit later meta-analytic comparison. In this sense, the proposed framework is not only an instructional model but also a reporting architecture for more precise research on audiovisual Russian-language education.

CONCLUSION

For classroom practice, the framework implies that teachers should plan audiovisual lessons by asking a sequence of cognitive questions. What information must learners hold simultaneously? Which visual cues directly support speech, and which may distract attention? What prior lexical or cultural knowledge is assumed? How much inference is required? How autonomous is the expected response? These questions produce a more precise design rationale than labeling a video simply as “appropriate for B1” or “interesting for students.” Proficiency labels remain useful, but they do not replace analysis of the task itself.

The framework can also guide material development. When a task is too complex, simplification need not begin by replacing authentic media with artificial speech. Complexity can be reduced through segmentation, pre-training of a small set of indispensable lexical items, temporary caption support, visual cueing, reduced response options, or separation of comprehension and production stages. Conversely, a task can be made more demanding without changing the media by removing prompts, increasing inferential requirements, asking learners to reconcile conflicting cues, or requiring transfer to a new communicative situation. Such manipulations allow researchers to study the effects of task complexity more rigorously because the underlying input can remain constant.

The present study is conceptual and therefore does not claim empirical effectiveness. Its next stage should be experimental validation with Russian-language learners at different proficiency levels. Future studies can test the inter-rater reliability of the five complexity variables, examine whether progression across the four levels predicts comprehension and productive performance, and compare cluster-based calibrated task banks with conventional teacher-selected audiovisual activities. Measures of perceived mental effort, comprehension accuracy, response quality, and delayed retention would be particularly informative.

In conclusion, increasing the effectiveness of audiovisual Russian-language instruction requires more than expanding the quantity of multimedia content. The decisive methodological task is to regulate what learners are cognitively required to do with that content. The proposed four-level framework-recognition, guided integration, inferential reconstruction, and productive transfer-provides a systematic basis for that regulation. By combining task-level cognitive calibration with cluster-level resource sharing, teacher moderation, and common assessment, the model offers a defensible route toward differentiated, scalable, and methodologically transparent audiovisual instruction.

REFERENCES

1. Chen, O., Paas, F., & Sweller, J. (2023). A cognitive load theory approach to defining and measuring task complexity through element interactivity. Educational Psychology Review, 35, 63. https://doi.org/10.1007/s10648-023-09782-w

2. Jin, C., & Yan, J. (2025). The effects of task complexity and task sequencing on L2 performance: A systematic review. The Language Learning Journal, 53(1), 114–141. https://doi.org/10.1080/09571736.2024.2345911

3. Li, C., Wei, L., & Lu, X. (2024). Task complexity and L2 writing performance of young learners: Contributions of cognitive and affective factors. The Modern Language Journal, 108(3), 741–770. https://doi.org/10.1111/modl.12954

4. Mayer, R. E. (2021). Multimedia learning (3rd ed.). Cambridge University Press. https://doi.org/10.1017/9781316941355

5. Mayer, R. E. (2024). The past, present, and future of the cognitive theory of multimedia learning. Educational Psychology Review, 36, 8. https://doi.org/10.1007/s10648-023-09842-1

6. Montero Perez, M. (2022). Second or foreign language learning through watching audio-visual input and the role of on-screen text. Language Teaching, 55(2), 163–192. https://doi.org/10.1017/S0261444821000501

7. Muñoz, C. (2022). Audiovisual input in L2 learning. Language, Interaction and Acquisition, 13(1), 125–143. https://doi.org/10.1075/lia.22001.mun

8. Shin, S.-Y., & Lee, S. (2024). Effects of the type of audio-visual input and listenability on young L2 learners' listening comprehension. TESOL Quarterly, 58(4), 1580–1608. https://doi.org/10.1002/tesq.3298

9. Tan, S., Samad, A. A., & Ismail, L. (2022). Systematic literature review on audio-visual multimodal input in listening comprehension. Frontiers in Psychology, 13, 980133. https://doi.org/10.3389/fpsyg.2022.980133

10. Umarov, A. A. (2026). O‘quv klasteri sharoitida rus tilini o‘qitishda audiovizual vositalardan foydalanishning metodik imkoniyatlari. Ta’limda Istiqbolli Izlanishlar, 8(2), 170–174.