Interdisciplinary Description of Complex Systems 24(4), 415-429, 2026

ONE NARRATIVE, MULTIPLE MEDIA: LEARNING AND MEMORY ACROSS TEXT, AUDIO, VIDEO AND VIRTUAL REALITY

Bruno Matijašević[1,*] and Nebojša Lujanović[2]

[1]Catholic University of Croatia
Zagreb, Croatia

[2]University of Split, Department of Communicology
Split, Croatia

DOI: 10.7906/indecs.24.4.4
Regular article

Received: 13 November 2025.
Accepted: 14 May 2026.

ABSTRACT

This study examines how different media formats influence short-term memory across three educational levels. The aim was to determine whether the modality of content delivery-text, audio, video, or virtual reality (VR)-affects the recall of narrative information among primary school pupils, secondary school students, and university undergraduates. A narrative text was standardized and presented in four formats, and recall was measured using a 40-item factual checklist. The results demonstrate a significant interaction between educational level and modality. In primary school pupils, text, audio, and VR produced comparable recall, while video led to notably lower performance. Among secondary school students, text and video yielded the highest recall, followed by audio, whereas VR produced the weakest results. In contrast, VR was the most effective modality in the university cohort, outperforming video, audio, and text, the latter being the least effective. These findings indicate a developmental shift in the cognitive processing of media. Younger learners appear to benefit from simpler, less visually dynamic media, while older learners show increased capacity to integrate and retain information from immersive VR environments. Overall, the results challenge the assumption that traditional text formats are consistently the most effective for short-term memory and highlight the pedagogical potential of VR, particularly in higher education. Future research should test these findings on larger samples, examine long-term retention, and explore cognitive load mechanisms underlying modality differences.

KEY WORDS

learning environment, educational material, short-term memory, virtual reality, quasi-experiment

CLASSIFICATION

JEL: I21, J13

*Corresponding author, η: bruno.matijasevic@unicath.hr; -; -

INTRODUCTION

Modern educational environments increasingly adopt multiple modes of content delivery-such as text, audio, video, and immersive virtual reality (VR) – as tools to enhance learning. Research in multimedia learning theory suggests that modality (i.e., the medium through which information is presented) significantly influences comprehension, retention, and cognitive load [1, 2]. For example, the modality effect (when auditory-only or dual-channel formats outperform visual-only ones) has been demonstrated in many contexts, though recent work indicates that such effects may reverse under certain conditions, such as within virtual reality environments [3]. Learning effectiveness is also moderated by developmental stage. Studies show that younger learners may benefit more from simpler, less cognitively demanding formats (such as text or audio), whereas older or more advanced learners often gain greater advantage from richer modalities, including video and immersive media, especially when these formats are designed to leverage dual processing channels or reduce cognitive overload. For instance, Radianti et al. [4] in their systematic review of immersive VR applications identify design features and learner characteristics (including age and educational level) as critical moderators of VR effectiveness. A complementary perspective is offered by Manfred Spitzer, who has argued extensively that digital media may impose cognitive demands that exceed the developmental capacity of younger learners. In his book *Learning: The Human Brain and the School for Life*, Spitzer [5] highlights that attention and memory resources are limited, particularly in children, and that complex multimodal input can easily overwhelm rather than enhance learning. In a related OECD report (*Brain Research and Learning over the Life Cycle*; Spitzer [6], he emphasizes that effective educational design must align with the maturational state of the learner's brain, warning that overly stimulating digital formats may hinder rather than foster knowledge acquisition in early stages of development. These insights provide a theoretical rationale for expecting different outcomes of modality effectiveness across educational levels, with immersive formats such as virtual reality potentially advantageous only once cognitive control and attentional capacities are more fully developed. However, despite growing interest in virtual reality and immersive learning media, relatively few studies systematically compare multiple modalities (text, audio, video, VR) within the same sample across different educational levels. As Liberman et al. [3] argue, while many studies apply modality principles, the specific performance of VR compared to more traditional modes is under-examined, especially in younger learners. This gap is particularly noteworthy given that immersive media may impose both enhanced sensory engagement and greater demands on cognitive resources, potentially resulting in differential effects depending on learners' age or educational background.

The present study aims to address this gap. Specifically, it examines how four modalities – text, audio, video, and VR – affect information retention in learners at three educational levels (primary school, secondary school, university). By comparing the modalities within and across these educational levels, this research seeks to uncover (1) whether modality effects vary by educational level, and (2) whether immersive formats such as VR offer advantages or disadvantages relative to more traditional media, especially as learners mature.

CURRENT INSIGHTS INTO THE EFFECT OF READING USING CONTEMPORARY MEDIA

The increasing prevalence of digital technology in reading in general, and in education specifically, necessitates thorough investigation into the effects of screen reading on various aspects of literacy. Numerous questions arise, including justified concerns regarding whether students retain information better when texts are presented on screens or on paper, whether

more intensive screen usage leads to poorer memory retention, and whether such influences extend to other reading components such as text comprehension, analysis, critical evaluation, and similar skills. Research should clarify these uncertainties to enable both educators and students to understand the impacts of technology and to mitigate negative consequences while highlighting its benefits.

Since 2010, a substantial body of research has addressed these issues, and this chapter aims to present selected studies primarily to illustrate the conflicting results and the careful reservations scholars make in interpreting them. This complexity arises not only from the intricate nature of the reading process itself but also from the multitude of variables that each study only partially addresses. On one hand, research indicates that reading on screens may lead to reduced comprehension or more superficial processing compared to reading on paper. However, studies with opposing conclusions must also be acknowledged [7]. A prominent meta-analysis examined studies from 2000 to 2017, confirming a significant advantage of paper-based reading over digital reading; this advantage reportedly increased over the review period. Specifically, this study depicted an "inferiority of screens", whereby digital texts resulted in poorer comprehension compared to printed texts [8]. The study introduced the "Shallowing Hypothesis", suggesting that the rapid interactions typical of digital media hinder readers' ability to engage in demanding cognitive tasks that require sustained attention, such as deep comprehension.

Neurophysiological investigations conducted over the past fifteen years, including participants whose reading skills developed under the influence of digital media, support these findings. One such study suggests that digital reading fosters considerably more superficial engagement with the text compared to printed reading [9]. While these examples could lead to the simple conclusion that digital media negatively affect text comprehension, such a stance over-simplifies the issue by ignoring numerous influencing variables in the reading process.

For instance, some studies demonstrate significantly better comprehension outcomes when reading from screens as opposed to printed paper. These contradictory findings highlight the danger of oversimplified conclusions that neglect individual differences in reading habits, learning styles, or familiarity with specific digital devices. It is also important to emphasize that electronic devices have the potential to offer engaging and enriched reading experiences through interactivity, such as hyperlinks and multimedia features [10].

Four selected studies summarize potential variations in interpretations and approaches to this problem. Some strongly favor paper over screens, others point out omitted variables, and some provide contradictory results. Although most focus predominantly on reading comprehension-the primary focus of this study being short-term memory-relevant conclusions can be drawn for memory as well. If research suggests more superficial text processing on screens compared to paper, it can be anticipated that this would influence the differentiation between short-term and long-term memory, given that deep comprehension is necessary for long-term retention. However, several questions arise: What is the nature of the text and its genre? What are the characteristics of the participants in terms of experience and abilities? Finally, how does this relationship appear in 2025, considering that six years have elapsed since the most recent referenced study-a substantial interval amid rapidly progressing changes?

PRELIMINARY CONSIDERATIONS BEFORE THE STUDY

The research presented thus far illustrates the inherent complexity of the reading process and how this complexity is amplified when reading occurs in digital environments. As digital technology becomes increasingly integrated into daily life and educational contexts, a central question emerges: *How does the medium of content presentation affect learners' immediate factual recall?* The underlying premise is that various media exert distinct influences on

specific aspects of the learning process, including recall, creating broad opportunities for examining both the nature and direction-positive or negative-of the effects associated with contemporary digital media.

Given the multifaceted nature of reading, this study focuses specifically on short-term memory as one component of the reading process. Memory in general represents a stage closely linked to the development of reading proficiency and varies considerably across individuals. As summarized by Kolić-Vehovec [11], the reading process begins with word recognition, proceeds through text comprehension-entailing meaning construction, information integration, inference-making, and the use of prior knowledge-and then advances to short- and long-term memory, ultimately culminating in metacognitive awareness. Although this overview simplifies the process, it clarifies that memory is only one of several possible reading outcomes and that others remain equally significant areas of inquiry.

It is also essential to acknowledge that there is no universal or uniform reader. Reading outcomes depend heavily on the reader's prior experiences, knowledge, and personal engagement, and thus adapt to individual conditions [12]. Consequently, research findings in this field refer to general tendencies in the development of reading skills rather than to a single standardized reading profile. More nuanced examinations of memory should therefore consider differences among reader types, including the possibility that sampled readers may differ in their orientation toward emotional engagement. Furthermore, variation in reading proficiency-from struggling to skilled readers-must be taken into account. Research in Croatia [13] has demonstrated clear differences in reading quality related to motivation, interests, cognitive capacity, vocabulary size, and prior knowledge. In alignment with these considerations, the present study is designed around a theoretically universal model of the reader and reading process, assuming a certain developmental level of reading skill. However, the study's relatively small sample size ($N = 48$) necessitates caution: the results should be interpreted as indicative rather than conclusive. The findings will serve as a theoretical basis for identifying trends and processes related to reading and memory in digital contexts, offering a foundation for continued research and discussion.

This perspective aligns with the work of German educator Hans E. Giehrl, who identified several reader types but cautioned against defining rigid categories. He describes the *functional-pragmatic reader*, focused on acquiring only essential information; the *emotional-fantastic reader*, oriented toward affective engagement and imaginative experience; and the *rational-intellectual reader*, who seeks to acquire knowledge and establish logical connections to form broader conceptual understanding [14]. These distinctions further support the importance of considering reader variability when examining the influence of different reading media.

MATERIALS AND METHODS

The goal of this study is to determine how different media-text, audio, video, and virtual reality-affect learners' recall of educational content at the primary, secondary, and university levels, in order to assess which media are pedagogically most appropriate for different developmental stages. This study was conducted across three educational levels: primary school, secondary school, and university. At each level, one intact class or cohort of students participated. The primary school sample consisted of pupils with a mean age of 11,6 years ($N = 16$), the secondary school sample comprised students with a mean age of 16,5 years ($N = 16$), and the higher education sample consisted of undergraduate students with a mean age of 18,5 years ($N = 16$). Secondary school participants attended a general education (gymnasium) program, whereas university participants were undergraduate students enrolled in a social sciences and humanities study program. Participation was voluntary, and informed consent was obtained from all

participants; for minors, consent was also secured from parents or legal guardians. The study adhered to the ethical principles of the Declaration of Helsinki and received approval from the relevant institutional review board. All testing sessions were conducted during regular school or university hours, in comparable classroom settings, under the same instructions and time constraints across all groups. Data collection took place in May 2025.

The research employed a quasi-experimental design with four conditions embedded within each educational level. In each institution (elementary school, highschool and university), participants were assigned to one of four groups, with each group exposed to the same informational content but presented through a different medium. The first group received the information as a printed transcript (text), the second group as an audio-only recording resembling a radio broadcast, the third group through a video presentation, and the fourth group in an immersive VR environment. This design allowed for systematic comparisons both within each educational cohort and across the three educational levels.

The stimulus material consisted of a standardized informational text, adapted into the three presentation formats. The audio versions were recorded in a professional studio to ensure clarity and uniform delivery. For the video version, the identical script was paired with synchronized imagery and narration. In the VR condition, the same content was embedded in an interactive environment that enabled participants to navigate virtually while receiving the information. To ensure comparability, the duration of exposure was held constant across all modalities, lasting 7,37 minutes. The procedure followed the same steps in each educational context. At each educational level, the total sample consisted of 16 participants. These participants were divided into four modality groups (text, audio, video, VR), with four participants assigned to each modality within each educational level. Thus, the experimental design comprised 12 independent groups in total (3 educational levels × 4 modalities), each containing four participants, resulting in a total sample size of $N = 48$.

Data were analyzed using SPSS (version 29.0). Descriptive statistics were first computed for all variables. Analyses of variance (ANOVA) were then performed to compare knowledge test scores and questionnaire responses across the four groups within each educational level. Within each educational level, modality served as a between-subjects factor, with four participants per modality. Importantly, analyses focused on patterns across modalities and educational levels rather than on isolated comparisons between individual groups of four participants. Post hoc comparisons were carried out using Tukey's HSD where appropriate. In addition, a two-way ANOVA was conducted to examine the interaction between educational level (primary, secondary, university) and medium of presentation (text, audio, video, VR). Effect sizes were reported using η^2, and statistical significance was set at $p < 0,05$.

RESEARCH DESIGN

This research was conducted as a controlled quasi-experiment using a between-subjects factorial design, with modality and educational level as the two independent variables. Such a structure allowed for systematic examination of the interaction between media format and developmental stage. The study employed a quasi-experimental between-subjects design, as intact school classes and a university cohort were used and participants could not be randomly assigned to the four modality conditions.

Since the aim of the study was to assess learners' immediate factual recall of narrative content delivered through different media, the measurable elements consisted of factual information integral to the text, which could be clearly and unambiguously identified and extracted to form a separate checklist. This checklist represented the maximum possible score, thereby establishing an exact quantitative component and producing measurable and comparable results. According to the authors' perspective, the most suitable text type for this evaluation is narrative text.

Accordingly, the starting point was the short film *Disident* directed by Ivan Mihaljević, with a duration of 7 minutes and 37 seconds.

The film portrays the life story of Hrvoje Gospočić, a formerly promising journalist and photo-reporter who became a victim of unfortunate circumstances and manipulation by connected individuals-incidents emblematic of the failings of liberal capitalism-that left him homeless. The autobiographical narration is structured around a linear and concise depiction of the protagonist's professional and emotional decline, delivered in a confessional tone. With minimal emotional fluctuation and predominantly objective tenor, the first-person narrator recounts his story through several chronologically arranged episodes-from career successes, economic setbacks, to his current social, emotional, and housing situation. The facts comprising these episodes were extracted into a concise control list, with each fact representing one point. In total, the list contained 40 facts or a maximum of 40 points. This system established the scoring method for what participants were to recall at the study's conclusion. It should be noted that the control list was not designed as a standardised psychometric instrument, but rather as a content-based checklist derived directly from the narrative material. Each item corresponds to a factual element explicitly presented in the source material, and the same list was used across all modalities as an objective scoring key for immediate factual recall. As such, formal validation procedures typically applied to attitudinal or psychological scales were not applicable.

The study was designed around the idea of comparing the effects of reading via different media formats. Therefore, the narrative content of the video was transformed into other media: the film's transcript was printed on paper; the same text was recorded as an audio file; and the video material was converted for consumption via VR glasses, providing a simulated virtual experience. In this manner, the identical narrative was presented to participants in four different formats or media: VR, video, audio, and text. After engaging with their assigned medium, participants completed a questionnaire in which they indicated which specific content from the presentation they remembered. Their answers were scored according to the control list and the results compared. Immediate factual recall was selected as the outcome measure because it represents an initial stage in the learning process, serving as a prerequisite for further comprehension, integration, and long-term retention of educational content.

A priori power analysis was conducted using G*Power. With the selected statistical test, expected medium effect size, $\alpha = 0,05$, and desired power of 0,80, the analysis indicated that a minimum total sample size of 45 participants was required. The final sample of 48 participants therefore met the threshold suggested by the power analysis. However, because the design includes multiple experimental conditions and the study is quasi-experimental with small cell sizes, the results should still be interpreted as preliminary and exploratory rather than fully conclusive. Participants included primary school pupils, secondary school students, and university undergraduates. Sixteen participants were selected from each institution, with four participants assigned to each of the four media types. Participant allocation was balanced according to abilities (grades) [1]. Immediately after exposure, participants completed the surveys, yielding a total of 48 completed questionnaires, whose results are presented graphically. Although each modality group consisted of four participants within each educational level, the total sample comprised 48 participants (3 educational levels × 4 modalities × 4 participants). While the number of participants per cell was limited, the overall sample size met the minimum requirements indicated by the a priori power analysis. Nevertheless, due to small cell sizes, the statistical analyses should be interpreted as exploratory rather than confirmatory.

RESEARCH RESULTS

When averaging across all educational levels, VR produced the highest overall recall (M = 4,83), followed by audio (M = 4,33), text (M = 4,25), and video (M = 4,08). Thus, while VR was not uniformly superior across cohorts, it emerged as the most effective medium overall. Conversely,

video produced the lowest mean recall in the aggregate, despite relatively strong performance among secondary school and university students. A one-way analysis of variance (ANOVA) was conducted to examine potential differences in recall across the four modalities. Given that only group means were available rather than individual-level data, this analysis must be interpreted with caution. The results indicated no statistically significant differences between modalities, $F(3, 8) = 0,06$, $p = 0,98$. While this finding suggests broad comparability between presentation formats when averaged across educational levels, it should not be considered definitive evidence due to the absence of variance estimates derived from raw data. Overall, the descriptive results reveal meaningful differences in modality effectiveness depending on the educational level, with VR being particularly advantageous in higher education, text and video proving more effective in secondary school, and text, audio, and VR performing similarly in primary school. Mean recall scores for each modality are presented across the three educational levels, Table 1. In the primary school sample, participants exposed to text (M = 3,00), audio (M = 3,00), and VR (M = 3,00) achieved identical average scores, whereas those exposed to video demonstrated substantially lower recall (M = 1.25). In the secondary school sample, the highest recall was observed in the text condition (M = 5,50), followed by video (M = 5,25), audio (M = 3,75), and VR (M = 2,75). A different pattern emerged at the university level, where VR yielded the highest performance (M = 8,75), followed by audio (M = 6,25), video (M = 5,75), and text (M = 4,25). Given the quasi-experimental nature of the study and the limited number of participants per cell, statistical results are reported primarily to support descriptive trends rather than to serve as sole evidence for strong inferential claims.

Table 1. Descriptive statistics by education level and modality criterion.

Education	Text	Audio	Video	VR
Primary	3,00 (1,15) [2,28, 3,72]	2,70 (1,16) [1,98, 3,42]	1,30 (1,06) [0,64, 1,96]	3,00 (0,94) [2,42, 3,58]
Secondary	5,80 (0,79) [5,31, 6,29]	4,10 (0,99) [3,48, 4,72]	5,00 (0,94) [4,42, 5,58]	2,70 (0,67) [2,28, 3,12]
University	4,90 (1,37) [4,05, 5,75]	6,40 (0,70) [5,97, 6,83]	5,60 (1,26) [4,82, 6,38]	8,80 (1,14) [8,10, 9,50]

A two-way analysis of variance was conducted to examine the effects of education level and modality on recall performance. The analysis revealed a significant main effect of education level, $F(2, 108) = 143,36$, $p < 0,001$, indicating that participants' recall varied substantially across primary school, secondary school, and university. A significant main effect of modality was also observed, $F(3, 108) = 3,69$, $p = 0,014$, suggesting that the medium of content delivery influenced recall outcomes. Most importantly, a significant Education × Modality interaction emerged, $F(6, 108) = 22,85$, $p < 0,001$, demonstrating that the effectiveness of each modality was strongly dependent on the educational level of the participants.

Inspection of group means indicated that in primary school, text, audio, and VR resulted in comparable recall scores, while video yielded markedly lower performance. In secondary school, text and video facilitated superior recall compared to audio and VR, with VR showing the weakest performance at this level. By contrast, in the university cohort, VR produced the highest recall, outperforming video, audio, and text, the latter being the least effective in this group. This pattern underscores a developmental shift: whereas younger learners benefited most from text and audio-based modalities, older learners showed a pronounced advantage when engaging with immersive VR. Post hoc comparisons using Tukey's Honestly Significant Difference (HSD) further confirmed that VR recall scores were significantly higher than video when data were aggregated across education levels ($p < 0,05$). Other pairwise differences between modalities did not reach statistical significance at the aggregate level, although the interaction analysis suggests

that specific contrasts are meaningful within individual educational cohorts. Taken together, these findings highlight the importance of considering the interaction between learner age and instructional medium. The results suggest that immersive modalities such as VR may be particularly beneficial for older students, while more traditional formats such as text and video may remain more effective for younger learners.

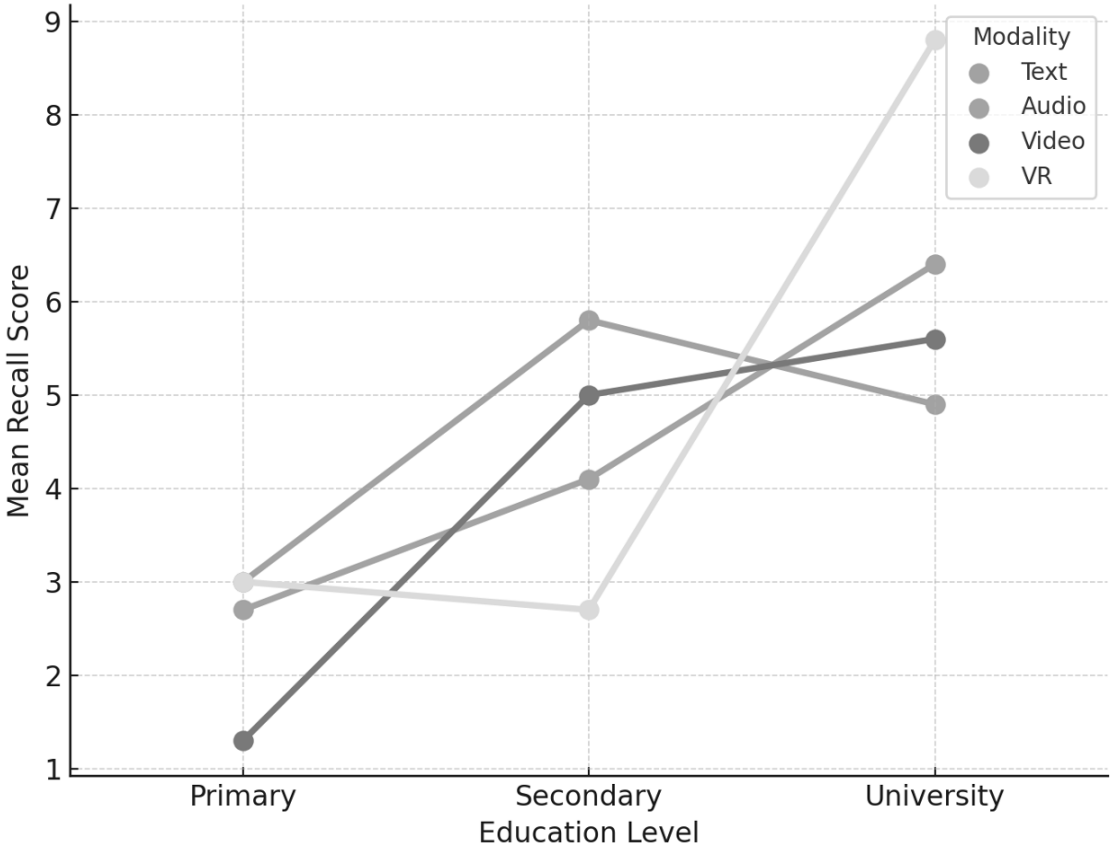

Figure 1. Visual presentation of interaction of Education Level and Modality Recall.

The two-way ANOVA revealed a significant interaction between education level and modality, $F(6, 108) = 22{,}85$, $p < 0{,}001$, indicating that the relative effectiveness of different modalities varied substantially across the three cohorts. To clarify this interaction, post hoc comparisons using Tukey's HSD were conducted within each educational level, Table 2. In the primary school sample, participants in the text, audio, and VR conditions performed comparably, whereas the video condition resulted in significantly lower recall than all other modalities ($p < 0{,}05$). This suggests that, at this developmental stage, auditory and textual formats, as well as immersive VR, are equally effective, while video is notably less effective for factual retention. In the secondary school sample, both text and video produced significantly higher recall scores compared to VR ($p < 0{,}01$). Moreover, text outperformed audio ($p < 0{,}05$), whereas no significant difference emerged between text and video. These results indicate that, for secondary school students, traditional text-based learning and audiovisual formats are superior to both audio-only and immersive VR. In the university cohort, VR yielded significantly higher recall than all other modalities ($p < 0{,}001$). In addition, both audio and video conditions led to significantly greater recall than text ($p < 0{,}05$), while audio and video did not differ from one another. Thus, in higher education, immersive VR proved to be the most effective medium, with traditional text trailing behind the other modalities. Taken together, these findings suggest a clear developmental shift in modality effectiveness. Whereas younger learners benefit equally from text, audio, and VR-with video being less effective-secondary school students achieve higher recall from text and video. At the university level, however, VR shows a distinct advantage, surpassing all other media formats.

Table 2. Post hoc comparisons (Tukey HSD).

Education	Group 1	Group 2	Mean.diff.	p-adj.
Primary	Audio	Text	0,3	0,9251
Primary	Audio	VR	0,3	0,9251
Primary	Audio	Video	–1,4	0,0314
Primary	Text	VR	0,0	1,0
Primary	Text	Video	–1,7	0,0064
Primary	VR	Video	–1,7	0,0064
Secondary	Audio	Text	1,7	0,0005
Secondary	Audio	VR	–1,4	0,0045
Secondary	Audio	Video	0,9	0,1075
Secondary	Text	VR	–3,1	0,0
Secondary	Text	Video	–0,8	0,1786
Secondary	VR	Video	2,3	0,0
University	Audio	Text	–1,5	0,0288
University	Audio	VR	2,4	0,0002
University	Audio	Video	–0,8	0,4133
University	Text	VR	3,9	0,0
University	Text	Video	0,7	0,5286
University	VR	Video	–3,2	0,0

Table 3 presents the relative ranking of the four modalities within each educational level. In the primary school sample, text, audio, and VR were equally effective, all occupying the top rank, whereas video consistently occupied the lowest position. This apperar toappears to align with the descriptive and post hoc analyses indicating that video was the least effective modality for younger learners. In the secondary school cohort, text achieved the highest rank, followed closely by video, while audio was ranked third and VR last. This pattern reinforces the finding that secondary school students benefitted most from traditional text-based and audiovisual formats, with VR being least effective. At the university level, VR clearly dominated, occupying the top rank and substantially outperforming the other modalities. Audio and video followed in second and third place respectively, while text consistently occupied the lowest rank. This shift suggests that immersive environments become more advantageous for advanced learners, while reliance on text diminishes. Overall, the ranking analysis complements the inferential results, highlighting a developmental trend: younger learners tend to perform equally well with text, audio, and VR but struggle with video, secondary school students favor text and video, whereas university students achieve maximal recall through VR.

The study conducted at the primary school level showed that participants achieved identical mean recall scores (M = 3,00) when the content was presented in text, audio, and VR formats. By contrast, the mean score for the video condition was notably lower than the other three media. These findings suggest that primary school pupils are more successful in retaining information presented through text, audio narration, and VR experiences, whereas video appears to be a less effective medium for this age group.

The results obtained from secondary school students indicate the highest mean recall score in the text condition, suggesting its strongest educational effectiveness within this cohort. This

Table 3. Modality ranks within each education level.

Education	Text	Audio	Video	VR
Primary	1	2	3	4
Secondary	1	3	2	4
University	4	2	3	1

was followed by the video condition, while the audio format resulted in moderately lower performance. Interestingly, the lowest mean recall score was recorded in the VR condition. These findings imply that, among secondary school learners, more traditional media formats – particularly text and video – are more effective for information retention than either audio or VR. Overall, the results suggest that secondary school students acquire educational content most efficiently when it is presented in textual or video form.

At the higher education level, the highest recall scores were observed in the VR condition, making it the most effective medium within this group. This was followed by the video and audio formats, respectively. The lowest recall results were found among students who engaged with the content in text form, indicating reduced effectiveness of the traditional reading-based approach in this age group. The results therefore suggest that university students learn and retain information more effectively through visually and interactively rich media, while text-based formats appear to be the least effective.

OVERALL VALUES AND INTERPRETATİON OF RESULTS

Across all three educational levels, substantial variation in recall performance was observed across media types. These differences may be related to developmental factors, prior educational experience, and media consumption habits, which should be explored in future research.

Comparing media effectiveness across age groups reveals notable developmental trends. The relative influence of text appears to decline with age, even though its absolute performance increases slightly (though at a slower rate than VR). The video medium shows the weakest performance in the youngest group,yet stabilizes at a comparatively high level among secondary school and university students, where it occupies a consistent second place. VR demonstrates a pronounced increase in effectiveness among university students, where it becomes the highest-performing medium. Audio maintains relatively stable absolute performance across age groups, but its relative position declines in comparison to other media.

When synthesizing these findings, it may be concluded that text gradually loses its dominant role in information retention over the course of educational development, while audio remains relatively important during earlier stages but does not retain this significance later. The youngest participants exhibit the greatest difficulty with video-based content, though its effectiveness increases and stabilizes with age. VR becomes increasingly effective as learners mature and appears to replace both text and audio for university-level learners-an important consideration when selecting media formats in instructional design.

Taken together, the results are both unexpected and promising for further investigation. This mostly refers to a certain advantage of the VR medium over the others in specific segments where a strong advantage of text had been expected. The remaining media show comparatively smaller differences, with a slight advantage for the audio medium; however, this effect is insufficient to support strong conclusions at this stage. Nevertheless, this trend aligns with cultural theory perspectives that emphasize a contemporary shift toward oral and visually mediated communication, facilitated by emerging technologies.

The overall advantage of VR in short-term memory performance suggests the need to reconsider the traditional assumption that text is inherently the most effective medium for retention. Should this trend be confirmed by larger-scale studies, it may indicate that VR represents a highly valuable educational medium. For now, these findings-derived from a small sample-invite reflection on the pedagogical potential of VR and the necessity of adapting reading and learning processes to evolving technological environments.

DISCUSSION

The present study examined the effectiveness of different modalities of content delivery-text, audio, video, and VR – across three educational levels. The results revealed a strong interaction between education level and modality, indicating that the impact of media on learning is highly dependent on learners' developmental stage.

In primary school, recall performance did not differ significantly between text, audio, and VR, whereas video yielded substantially lower outcomes. This pattern suggests that younger learners may not benefit from visually rich formats such as video, which can impose unnecessary cognitive load on still-developing attentional and working memory systems. Theories of multimedia learning [1, 2] argue that younger learners have limited capacity for dual-channel processing, and as Spitzer [6] emphasized, multimodal stimulation that exceeds these limits may result in distraction rather than deeper processing. Our results are consistent with this perspective, showing that simpler modalities-particularly text and audio-can be as effective as, or more effective than, immersive or visually complex presentations in early learning stages. At the secondary school level, text and video emerged as the most effective modalities, while VR was least effective. This developmental shift is noteworthy: adolescents appear increasingly able to benefit from video, likely due to improved dual coding and integration of auditory and visual information [1]. Yet the weakness of VR in this group points to the challenges of immersion without sufficient scaffolding. As Radianti et al. [4] demonstrated, the design of VR applications for education requires careful balance: immersive features can enhance engagement, but without clear instructional structure they may overwhelm learners. This echoes Spitzer's [5] argument that instructional design must align with maturational capacities, and in adolescence, cognitive resources may still be insufficient to fully capitalize on immersive experiences. In contrast, the university cohort demonstrated a striking advantage for VR, which significantly outperformed all other modalities. Here, the immersive environment appears to have provided meaningful benefits, consistent with research showing that older learners are able to manage complex sensory environments and leverage them for deeper engagement and retention [3]. Recent reviews have further emphasized that immersive media in higher education are particularly effective in fostering critical thinking and higher-order skills when used purposefully within the curriculum [15]. The superiority of VR in our university sample aligns with these findings, suggesting that immersive learning environments become increasingly advantageous as learners' cognitive control and metacognitive regulation mature.

The pattern observed in primary school deserves closer examination, as it speaks directly to the cognitive ecology of younger learners. Three modalities – text, audio, and VR – produced identical mean recall scores (M = 3,00), while video alone fell substantially behind (M = 1,30). At first glance, this result is counterintuitive: video typically combines auditory and visual channels in a way that should facilitate dual coding [1]. Several explanations may account for the underperformance of video at this age. First, the video condition introduces a continuous stream of visual information that competes with the auditory narration for limited attentional resources. According to cognitive load theory [2], young learners with less developed executive control are particularly vulnerable to the split-attention effect, where simultaneous decoding of imagery and speech overwhelms working memory rather than enriching it. Second, video imposes pacing externally; the learner cannot slow down, re-read, or re-listen at will, unlike text or audio. For learners who are still consolidating reading fluency and listening comprehension skills, this loss of self-pacing may be especially consequential. Third, the comparable performance of VR with text and audio in this cohort, although seemingly surprising, may reflect the novelty effect documented in immersive media research [4]: immersion captures attention strongly enough to compensate for limited cognitive control, even

if it does not yet provide the integrative advantages observed in older learners. Taken together, these findings suggest that for primary school pupils the most pedagogically reliable modalities are those that allow either internal pacing (text, audio) or strong attentional capture (VR), whereas conventional video – despite its ubiquity in classrooms – may be the least suitable for short-term factual recall at this developmental stage.

The secondary school results present a complementary puzzle that is equally informative. Adolescents in this sample performed best with text (M = 5,80) and video (M = 5,00), while VR (M = 2,70) yielded the weakest recall – the opposite of the university pattern. This reversal is theoretically meaningful and cannot be reduced to a simple developmental progression in which immersive media become uniformly more useful with age. Several mechanisms may underlie the secondary school disadvantage for VR. First, adolescence is a period of substantial reorganization of the prefrontal cortex and of the attentional networks involved in cognitive control [5, 6]; immersive environments place demands precisely on these still-maturing systems, which may result in attentional capture without effective integration. Second, secondary school students typically have well-established study habits centered on text and audiovisual classroom material, but limited structured experience with VR as a learning medium; in the absence of metacognitive strategies adapted to immersion, the novelty of VR may translate into distraction rather than engagement. Third, the strong performance of text in this cohort suggests that secondary school students have already acquired the decoding fluency necessary to extract dense factual information from written material efficiently, while video supplements this with synchronized visual cues that reinforce – rather than compete with – the narrative. The implication is that VR is not categorically superior at any age; rather, its effectiveness depends on a match between the cognitive demands it imposes and the learner's capacity to regulate attention within the immersive frame. For mid-adolescent learners, this match appears not yet to have been achieved, which has direct implications for instructional design: introducing VR into secondary education without explicit scaffolding for attention management and content integration may yield disappointing learning outcomes despite high engagement.

A complementary explanation for the strong VR performance among university students may be drawn from a distributed and embodied account of reading. Trasmundi et al. [19] argue that reading is not a purely internal decoding process but an embodied, multi-scalar activity in which the reader's body, the material affordances of the medium, and the surrounding environment jointly shape comprehension and retention. Within this framework, the sensorimotor engagement that VR demands – turning the head, attending to spatial cues, navigating a simulated environment – can be understood not as an extraneous distraction but as an additional channel through which information becomes anchored in experience. This perspective also resonates with Aarseth's [20] notion of ergodic texts, in which "nontrivial effort" is required of the user to traverse the work. VR is, in this sense, ergodic in a way that printed text, audio, and video are not: the learner must actively perform the experience rather than passively receive it. Such active traversal may help explain why immersive VR produced the highest recall in the university cohort, where learners possess the cognitive control needed to convert sensorimotor engagement into structured retention rather than into cognitive overload.

When the results are considered in aggregate across all three levels, VR produced the highest mean performance overall. However, the strong Education × Modality interaction underscores that this overall trend is misleading: VR is not uniformly effective across developmental stages, and its benefits emerge most clearly in older learners. This aligns with recent empirical work demonstrating that the effects of immersion are conditional. For example, Zhang, Luo, Peng, and Han [16] found that VR yielded significantly larger learning effects than videoconferencing, particularly in higher education contexts, while Lu, Zhang and Lee [18] showed that multimodal environments reduce cognitive load and enhance interaction when well-structured. Together with Spitzer's neurodidactic framework, which stresses the alignment of pedagogy with

learners' cognitive capacity, these findings suggest that the key to modality effectiveness lies in the match between medium, learner maturity, and instructional design.

LIMITATIONS AND FUTURE DIRECTIONS

Several limitations of the present study should be acknowledged when interpreting the findings. First, the sample size within each experimental group was very small, with only a few participants per modality at each educational level. This limited the statistical power of the analyses and reduced the reliability of between-group comparisons.

Second, the statistical analyses were conducted using aggregated group means rather than individual-level data. As a result, the ANOVA results should be interpreted with caution because variance estimates were not derived from raw observations. While these procedures provide an indicative overview of possible patterns, they cannot offer fully reliable inferential conclusions. Third, although all four modalities were based on the same narrative and were matched in duration, they differed inherently in perceptual richness, visual complexity, and immersive qualities. The VR condition, for example, naturally introduces a higher degree of immersion than text or audio, while the video condition contains dynamic visual elements absent in both. These differences make it impossible to attribute effects solely to the modality variable and introduce potential confounds related to sensory load and media characteristics. Fourth, the study relied exclusively on immediate recall of factual information as the outcome measure. This focus provides insight into short-term memory but does not capture deeper comprehension, long-term retention, or higher-order cognitive processing. Additionally, the study did not account for participants' prior experience with digital media or VR, which may have influenced performance, particularly in immersive conditions. Finally, the quasi-experiment was conducted in natural school environments, which may have introduced uncontrolled external factors such as noise, distractions, or variability in technical conditions. These contextual variables may have affected participants' engagement and recall.

Taken together, these limitations indicate that the present findings should be considered preliminary. Future research would benefit from larger and more balanced samples, access to individual-level data, additional cognitive measures, and more tightly controlled comparisons of media modalities. Additionally, the choice of narrative content may represent a further limitation. The selected material – a documentary about a socially sensitive personal story – may have elicited varying emotional responses among participants, which could have differentially influenced attention and recall across age groups. Future studies should consider using emotionally neutral or thematically distant content to minimise potential confounding effects of emotional engagement with the topic.

Future studies should explore three directions. First, replication with larger, more diverse samples and raw data will allow finer-grained analyses of individual differences. Second, longitudinal designs assessing delayed recall are needed to test the durability of modality effects. Third, examining cognitive load using behavioral, self-report, and physiological indicators would clarify when immersive or complex media support learning and when they hinder it. Recent evidence suggests that design features such as signaling, pacing, and segmentation [17, 18] can mitigate overload in multimedia and VR learning. Further, investigating learner characteristics such as working memory, spatial ability, and prior exposure to digital media may help identify for whom immersive formats like VR are most effective.

CONCLUSION

This study shows that the effectiveness of learning modalities is not uniform but shifts with educational development. While primary school learners benefit equally from text and audio, and struggle with video, secondary students gain most from text and video, and university

students excel in immersive VR. The results should be interpreted as indicative rather than conclusive, providing a basis for further controlled studies with larger samples. These findings highlight a developmental trajectory in which simpler formats support early learning, whereas richer and more complex modalities become advantageous as cognitive capacity matures. The results extend multimedia learning theory and neurodidactic perspectives by demonstrating that modality effectiveness is conditional on both cognitive readiness and instructional design. Practically, they caution against uncritical adoption of advanced technologies in early education while pointing to the transformative potential of VR in higher education. By aligning media with learners' developmental stage, educators and policymakers can maximize both engagement and retention, ensuring that technological innovation truly serves the goals of learning. At the same time, the study underscores that technology itself is not a guarantee of improved outcomes. Rather, it is the thoughtful integration of media into pedagogy-sensitive to learners' developmental needs-that determines whether digital tools function as aids or obstacles to learning. By situating modality within a developmental framework, this research contributes to both theory and practice, offering guidance for the effective use of traditional and emerging media in education.

REMARK

[1]In each of the three mentioned institutions, teachers/professors were instructed to pay attention to the structure of the sample. They created a sample consisting of four groups, ensuring that each group included a comparable number of students with low, average, and excellent grades. In this way, all groups (which received the narrative through different media) had an equivalent profile in terms of student quality or academic performance. Specifically, in the primary school sample (where students with low grades were not present), each modality group of four students included 2 with average grades and 2 with excellent grades. In both the secondary school and university samples, each modality group of four students included 1 with low grades, 2 with average grades, and 1 with excellent grades. This balanced allocation ensured that observed differences across modalities cannot be attributed to systematic differences in academic ability between groups.

REFERENCES

[1] Mayer, R.E., ed.: *The Cambridge Handbook of Multimedia Learning*. Cambridge University Press, 2005, http://dx.doi.org/10.1017/CBO9780511816819,

[2] Ginns, P.: *Meta-analysis of the modality effect*. Learning and Instruction **15**(4), 313-331, 2005, http://dx.doi.org/10.1016/j.learninstruc.2005.07.001,

[3] Liberman, L. and Dubovi, I.: *The effect of the modality principle to support learning with virtual reality: An eye-tracking and electrodermal activity study*. Journal of Computer Assisted Learning **39**(2), 547-557, 2023, http://dx.doi.org/10.1111/jcal.12763,

[4] Radianti, J.; Majchrzak, T.A.; Fromm, J. and Wohlgenannt, I.: *A systematic review of immersive virtual reality applications for higher education: Design elements, lessons learned, and research agenda*. Computers & Education **147**, No. 103778, 2020, http://dx.doi.org/10.1016/j.compedu.2019.103778,

[5] Spitzer, M.: *Learning: The Human Brain and the School for Life*. Elsevier Science, Amsterdam, 2007,

[6] Spitzer, M.: *Brain Research and Learning over the Life Cycle*. In: OECD: *Personalising Education*. OECD Publishing, pp.47-62, 2006, http://dx.doi.org/10.1787/9789264036604-en,

[7] Mangen, A.; Walgermo, B.R. and Brønnick, K.: *Reading linear texts on paper versus computer screen: Effects on reading comprehension*.
International Journal of Educational Research **58**, 61-68, 2013,
http://dx.doi.org/10.1016/j.ijer.2012.12.002,

[8] Delgado, P.; Vargas, C.; Ackerman, R. and Salmerón, L.: and: *Don't throw away your printed books: A meta-analysis on the effects of reading media on reading comprehension*.
Educational Research Review **25**, 23-38, 2018,
http://dx.doi.org/10.1016/j.edurev.2018.09.003,

[9] Froud, K.; Levinson, L.; Maddox, C. and Smith, P.: *Middle-schoolers' reading and processing depth in response to digital and print media: An N400 study*.
preprint bioRxiv, 2023,
http://dx.doi.org/10.1101/2023.08.30.553693,

[10] Tajuddin, E.S. and Mohamad, F.S.: *Paper versus screen: Impact on reading comprehension and speed*.
Indonesian Journal of Education Methods Development **3**(2), 2019,

[11] Kolić-Vehovec, S.: *Cognitive and metacognitive aspects of reading*. In Croatian.
In: Mićanović, M., ed.: *Proceedings of the 4th Symposium of Croatian Language Teachers Reading for School and Life*. In Croatian. Agencija za odgoj i obrazovanje, Zagreb, pp.23-32, 2013,

[12] Kuvač Levačić, K.: *Development and types of reading, typology of readers with regard to reading non-literary texts*. In Croatian.
In: Mićanović, M., ed.: *Proceedings of the 4th Symposium of Croatian Language Teachers Reading for School and Life*. In Croatian. Agencija za odgoj i obrazovanje, Zagreb, pp.13-22, 2013,

[13] Rončević Zubković, B.: *Reading Skill in Early Adolescence: Student Profiles*. In Croatian.
Društvena istraživanja **20**(1), 189-210, 2011,
http://dx.doi.org/10.5559/di.20.1.10,

[14] Kordigel, M.: *Development of reading, types of reading and typology of readers*. Part II. In Croatian.
Umjetnost i dijete 4, 175-193, 1991,

[15] Balalle, H.: *Learning beyond realities: Exploring virtual reality, augmented reality, and mixed reality in higher education-A systematic literature review*.
Discover Education **4**, No. 151, 2025,
http://dx.doi.org/10.1007/s44217-025-00559-7,

[16] Zhang, Y.; Luo, H.; Peng, S. and Han, X.: *Virtual reality or videoconferencing for online learning? Evidence from comparative meta-analyses*.
Applied Sciences **15**(11), No. 6293, 2025,
http://dx.doi.org/10.3390/app15116293,

[17] Bali, C.; Várkonyi, G.; Szabó, M. and Zsidó, A.N.: *The impact of visual cues on reducing cognitive load in interactive storybooks for children*.
Journal of Experimental Child Psychology **260**, No. 106320, 2025,
http://dx.doi.org/10.1016/j.learninstruc.2025.101126,

[18] Lu, M. and Hu, Z.: *Leveraging Multimodal Information for Web Front-End Development Instruction: Analyzing Effects on Cognitive Behavior, Interaction, and Persistent Learning*.
Information **16**(9), No. 734, 2025,
http://dx.doi.org/10.3390/info16090734,

[19] Trasmundi, S.B.; Kokkola, L.; Schilhab, T. and Mangen, A.: *A distributed perspective on reading: implications for education*.
Language Sciences **84**, No. 101367, 2021,
http://dx.doi.org/10.1016/j.langsci.2021.101367,

[20] Aarseth, E.J.: *Cybertext: Perspectives on Ergodic Literature*.
Johns Hopkins University Press, Baltimore, 1997,
http://dx.doi.org/10.56021/9780801855788.