Beyond Reality: Virtual Storytellers and the New Age of Visual Literacy

Safa AlOthali[1], Maha Alhabbash[1], Najah Al Mohammedi[1*], Negmeldin Alsheikh[1], Hana Omar[1]
[1]United Arab Emirates University, United Arab Emirates

Abstract This case study investigates the impact of Virtual Reality (VR) visual storytelling on reading-related engagement, comprehension, and retelling performance, compared with television (TV). The study focuses on fourth-grade bilingual Arab students' engagement with the same silent visual story presented in both VR and TV modes. Adopting a mixed-methods design, quantitative data were collected through an observational checklist and a self-report scale, followed by qualitative video analysis guided by sequential analysis and reflexivity. Quantitative findings indicated that VR outperformed TV across all analytical categories. Self-report results showed that students rated VR higher in vividness, immersion, character identification, and perceived interactivity. Qualitative video analysis identified three themes: immersive environments that support sustained engagement and active language use; multisensory engagement that facilitates comprehension and narrative recall; and emotional engagement that shapes participation across visual modes. Video evidence demonstrated embodied interaction with the story, coherent narrative reconstruction, translanguaging practices during retelling, and individualized interpretations of visual events. The findings provide evidence that VR visual storytelling enhances visual literacy engagement, narrative coherence, and emotional involvement more effectively than TV, particularly in bilingual learning contexts.

Keywords: *Virtual reality (VR), TV, Visual literacy, Silent story retelling, Elementary stage, Language*

[*]Corresponding Author:
Najah Al Mohammedi
201370306@uaeu.ac.ae

Received: November 9, 2025
Revised: January 5, 2026
Accepted: January 21, 2026
First Published: February 9, 2026

© 2026 AlOthali, Alhabbash, Al Mohammedi, Al Sheikh, and Omar.
This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY).

1. Introduction

In the late twentieth century, literacy had already diverged from its earlier forms, a shift that has intensified as contemporary communication extends beyond written language to include images, colour, sound, and moving visuals (Christie, 1997; Emery et al., 2019). Eisner (1997) defines literacy as "a way of conveying meaning, through and recovering meaning from the form of representation in which it appears" (p. 353), highlighting the role of multiple representational modes in meaning-making. In response, literacy has expanded to encompass diverse forms reflective of modern communication, with visual literacy emerging as a central component of the twenty-first-century literacy

https://doi.org/10.22034/ijscl.2026.2082623.4336

(Kress, 2003; Serafini & Gee, 2017; Walsh, 2021). Visual literacy has deep historical roots, from prehistoric cave paintings to contemporary visual media, underscoring the enduring role of visuality in human experience (Barnard, 2001; Mirzoeff, 2015). This "visual turn" reflects a cultural landscape increasingly shaped by visual texts that function as primary carriers of meaning rather than mere illustrations (Sturken & Cartwright, 2001).

Although visual literacy has been theorized as a multidimensional construct involving perception, interpretation, critical thinking, and production (Brumberger, 2011), its application within specific visual storytelling modes remains unevenly explored. Visual storytelling uses visual data to construct narratives that convey grounded events and subjective experience (Huang et al., 2016). Children's stories, in particular, are visually rich texts that communicate cultural values and role models aligned with children's development (Olshansky, 2008). While research on children's responses to visual texts has expanded (Farrar et al., 2024), visuals are often treated as supplementary rather than as structured systems of meaning. Moreover, limited attention has been paid to how different visual modes, such as television and emerging immersive technologies, shape children's engagement and narrative responses.

Recent advances in visualization technologies, including Virtual Reality (VR), have introduced immersive and multisensory storytelling experiences that differ qualitatively from traditional screen-based media (Wu et al., 2020). Although VR has been shown to enhance engagement and experiential learning (Papanastasiou et al., 2019), few studies have systematically compared VR with conventional visual media, such as television (TV), in the context of children's storytelling and literacy development.

This gap is particularly salient in Arab and bilingual educational contexts. In the United Arab Emirates (UAE), literacy increasingly involves the interpretation of diverse visual texts encountered in everyday life (Al Mohammedi, 2022). Despite this shift, reading outcomes remain below international benchmarks; for example, UAE students scored an average of 417 in reading in the Program for International Student Assessment (PISA) 2022, compared to the OECD average of 476 (OECD, 2023). As schools and families face challenges in sustaining reading habits, there is a growing need for pedagogical approaches that support learners' ability to interpret complex visual texts across modes (Rowsell et al., 2012).

Addressing this need, the present study investigates how different visual storytelling modes, VR and TV, support visual literacy and storytelling among fourth-grade bilingual students. Grounded in multimodal literacy theory (Kress, 2003; Kress & van Leeuwen, 2021), the study focuses on students' engagement with silent visual stories, particularly story retelling. By comparing students' retelling performances and responses across VR and TV, the study examines how immersion, interactivity, and narrative structure influence comprehension and storytelling, while also foregrounding students' perceptions and subjective experiences of visual media. This study aims to answer these questions:

1. Is there a significant difference in fourth graders' retelling scores when employing VR or TV for silent visual storytelling?
2. Do fourth graders perceive VR and TV as effective tools for reading and retelling silent stories?
3. How do fourth graders respond to silent visual storytelling using VR or TV?

2. Theoretical Framework

This study is based on Multimodal Literacy Theory, complemented by Visual Literacy Theory. These frameworks provide a solid foundation for examining how different visual modes, such as VR and TV, aid in retelling in literacy learning, while addressing the interaction of semiotic modes and the role of visual representation in meaning-making.

2.1. Multimodal Literacy Theory

Multimodal literacy theory conceptualizes meaning-making as the interaction of multiple semiotic modes, including visual, linguistic, spatial, auditory, and interactive elements (Kress, 2003; Kress & van Leeuwen, 2020). Literacy, therefore, involves actively designing and interpreting meaning across modes rather than passively consuming print. Gee's (2014) multimodal principle emphasizes that comprehension emerges through the coordinated use of images, actions, symbols, and interaction, a

process that is central in silent visual storytelling, where meaning depends on visual sequencing, spatial cues, emotional expression, and inferential reasoning (Gee, 2014; Kress, 2003).

From this perspective, VR constitutes a highly multimodal environment characterized by immersive visuals, embodied spatial navigation, and interactivity, whereas TV offers a more constrained multimodal experience with fixed viewpoints. Empirical studies show that immersive VR can support greater engagement, deeper cognitive processing, and more coherent mental model construction than non-immersive video formats when cognitive load is managed appropriately (e.g., Makransky & Petersen, 2021). VR-based learning has also been shown to enhance literacy outcomes through interactive visual exploration and multimodal meaning-making (Somadayo & Jamil, 2025), with narrative studies demonstrating more spatially organized and coherent retellings in VR than in traditional screen-based viewing (Dooley, 2024; Ruszev, 2021). In this study, multimodal literacy theory informed data collection by framing VR and TV as environments with differing multimodal affordances, and guided data analysis by interpreting retelling performance as evidence of learners' integration of visual, spatial, and narrative modes.

2.2. Visual Literacy Theory

Visual literacy theory focuses on how images communicate meaning, emphasizing that visuals are culturally and socially constructed texts rather than neutral representations (Barthes, 1985; Berger, 2015). Visual literacy involves decoding symbols, interpreting visual conventions and emotional cues, and inferring implicit meanings, positioning learners as active meaning-makers (Brumberger, 2011). Contemporary scholarship highlights the growing role of images in literacy and cognition, arguing that modern literacy requires navigating visually rich texts and developing visual comprehension skills (Metros, 2008; Serafini, 2012; Tillmann, 2012). Evidence from wordless books further demonstrates that image-based texts can enhance children's interpretive and comprehension abilities (Zhang et al., 2022).

Within this framework, VR provides an immersive visual environment characterized by spatial depth, first-person perspective, and embodied interaction, while TV offers a more static, observational visual experience. Research indicates that immersive visual environments enhance attention, emotional engagement, and narrative inference through increased spatial presence and visual coherence (Daghriri et al., 2024; Makransky & Petersen, 2021). In this study, visual literacy theory guided data collection by focusing on learners' interpretation of visual details, emotional cues, and narrative sequencing, and informed data analysis by shaping the evaluation of retelling quality and students' perceptions as indicators of visual comprehension across VR and TV.

Together, multimodal and visual literacy theories provided the analytical framework for comparing VR and TV in silent visual storytelling tasks. These theories guided the selection of story retelling and student responses as primary data sources and informed the interpretation of differences in retelling performance, narrative organization, and perceptions as reflections of how immersive and non-immersive media affordances support meaning-making, directly addressing the study's purpose and research questions.

2.3. Visual Literacy, Silent Storytelling, and Immersive Media in Children's Reading Comprehension

Visual literacy is central to children's reading comprehension, particularly in silent visual narratives where meaning is constructed through images rather than written language. It involves interpreting visual representations, spatial relations, and emotional cues through inferencing and narrative sequencing (Serafini, 2014). Such texts reduce linguistic barriers while engaging higher-order cognitive processes, including prediction, synthesis, and interpretation (Arizpe & Styles, 2015). For elementary learners, especially fourth graders, visual narratives enhance engagement and comprehension when attention is directed to visual and emotional details, though explicit instructional support remains necessary (Pantaleo, 2025; Serafini, 2024). While this perspective emphasizes comprehension processes at the learner level, understanding how visual narratives function as meaning-making systems requires a broader semiotic and multimodal lens. From a semiotic and multimodal perspective, visual texts function as complex meaning-making systems shaped by design, context, and function (Liu, 2013). Visual literacy is therefore a multidimensional competence encompassing perceptual sensitivity, critical analysis, aesthetic awareness, and creative production, enabling learners to recognize intention,

bias, and cultural context in visual media (Avgerinou & Pettersson, 2011; Ekoç, 2021). This view aligns with the contemporary pictorial turn, which positions images as dominant communicative resources across digital environments (Mirzoeff, 2015; Pauwels, 2020). Meaning emerges through the interaction of visual, linguistic, spatial, and interactive modes rather than any single mode alone (Gee, 2014; Kress, 2003), highlighting the need for intentional instruction despite children's extensive exposure to visual media (Pratt & Searles, 2017; Zakime, 2017).

This multimodal understanding of visual literacy directly informs how comprehension can be examined in instructional contexts that rely exclusively on visual meaning-making. Silent visual storytelling offers a valuable context for examining reading comprehension, as learners must infer causality, organize events, and construct coherent narratives solely from visual cues (Arizpe & Styles, 2015; Serafini, 2014). Retelling is a well-established indicator of comprehension because it requires narrative reconstruction, inferencing, and sequencing rather than surface-level recall (Kendeou et al., 2020; Morrow, 1985; Paris & Paris, 2003). For elementary learners, retelling silent visual narratives provides a sensitive measure of comprehension quality and cognitive engagement (Pantaleo, 2025; Paris & Hamilton, 2014).

Building on this foundation, recent technological developments have introduced new modalities that reshape how visual narratives are experienced and comprehended. Advances in digital technology, particularly VR, extend visual storytelling by offering immersive, interactive environments characterized by presence and multisensory engagement (Makransky & Petersen, 2019; Radianti et al., 2020). Empirical research indicates that VR enhances emotional engagement, narrative immersion, and comprehension in literacy and language learning contexts (Al Mohammedi, 2022; Gadelha, 2023; Papanastasiou et al., 2019; Wu et al., 2020). However, despite these documented affordances, research comparing children's silent reading and retelling of visual narratives in immersive VR and traditional media such as television remains limited, motivating the present study (Ibourk & Al-Adwan, 2019; Peixoto et al., 2021; Sweidan & Darabkh, 2018).

3. Methodology

3.1. Research Design

This case study adopted a mixed-methods explanatory sequential design, combining quantitative and qualitative approaches to provide a deeper understanding of the phenomenon under investigation. Quantitative data were first collected using an observational checklist and a self-report scale, followed by qualitative data collected through video analysis. The rationale for employing a mixed-methods approach was to obtain more in-depth data (Creswell & Plano Clark, 2023) that would support a richer understanding of how visual literacy enhances retelling skills. By comparing students' retelling performance and responses across two visual modes, the study sought to identify how immersion, interactivity, and narrative structure influence children's comprehension and storytelling abilities.

3.2. Participants

Six fourth-grade students were purposively selected to conduct this study based on their willingness and availability. The recruitment process involved specific criteria to ensure a targeted sample (Creswell & Guetterman, 2024). All participants were fourth-grade bilingual Arab students, aged 8-10 years, with an equal gender distribution of three females and three males (see Table 1).

Table 1
Demographic Information

Participant	Grade	Age	Gender	Nationality
P1	Grade 4	8	Female	Emirati
P2	Grade 4	9	Male	Emirati
P3	Grade 4	9	Female	Jordanian
P4	Grade 4	10	Male	Sudanese
P5	Grade 4	8	Female	Egyptian
P6	Grade 4	10	Male	Emirati

The decision to limit the sample to six participants was intentional and methodologically driven. Given the exploratory nature of the study and the intensive data-collection methods employed, including individual engagement with two visual storytelling modes, story retelling tasks, observations, and video-recorded interactions, a small sample enabled a close, fine-grained analysis of students' meaning-making processes. Data collection continued until saturation was reached, as no substantively new patterns, responses, or interpretive categories emerged in students' storytelling, engagement, or perceptions across participants (Bryman, 2016; Creswell & Guetterman, 2024). At this point, increasing the number of participants was unlikely to yield additional analytical insight relevant to the study's aims.

The study was conducted in a mainstream school in the UAE that uses both Arabic and English as languages of instruction, reflecting a bilingual educational context. All participants had prior exposure to basic digital technologies commonly used in classroom settings; however, none had formal experience with VR-based storytelling. Ethical approval was obtained through the school, and informed consent procedures were followed, with parents informed that students' interactions would be documented through video recordings for research purposes (Creswell & Guetterman, 2024).

3.3. Instrumentation

Due to the nature of the study, data collection was carried out in two phases. During the quantitative phase, an observational checklist was employed while students engaged with the two visual presentation modes (TV and VR). Upon completion of each viewing session, students were asked to rate their experiences using a self-report scale. The qualitative phase involved video recordings of students interacting with both visual modes, providing additional insight into their behavioral responses and engagement throughout the storytelling process.

3.3.1. Observational Checklist

An observational checklist was used to document students' behavioral and expressive responses during engagement with the visual stories presented via TV and VR (Appendix A). The checklist focused on visual attention, bodily engagement, and storytelling performance and consisted of eight items: elaboration on oral expression, level of watching attainment, body posture and focus, facial expression, body language, interactive gestures, reaction to events, and gist recalling. Each item was rated on a 10-point scale, with higher scores indicating stronger behavioral presence.

The checklist was developed based on relevant literature to ensure content and construct validity. In addition, the checklist was reviewed and revised by an expert panel in literacy education and educational technology, further substantiating its content and construct validity. In terms of reliability, inter-rater reliability was considered by using three independent raters who evaluated students' performance separately under the TV and VR conditions, reducing observer bias and enhancing scoring consistency.

3.3.2. Self-Report Scale

A self-report scale was administered to capture students' perceptions and emotional engagement with the storytelling experience (Appendix B). The scale consisted of 10 literature-based statements addressing immersion, identification with characters, agency, enjoyment, and perceived realism. Students rated each statement on a 10-point Likert-type scale, with higher scores indicating greater agreement.

The self-report scale was developed from existing literature on immersive storytelling and learner engagement. To ensure content validity, it was reviewed by an expert panel that evaluated its relevance, clarity, and appropriateness for fourth-grade bilingual learners. Their feedback helped refine the items to accurately reflect the targeted constructs in an understandable way for the age group. This process also enhanced the scale's construct validity by aligning it with the study's theoretical framework. Designed to be accessible, the scale allows students to express their experiences intuitively. When used with the observational checklist, it provides valuable insights into students' internal perceptions and emotional responses, enhancing the overall validity of the data collection.

3.3.3. Video Recording

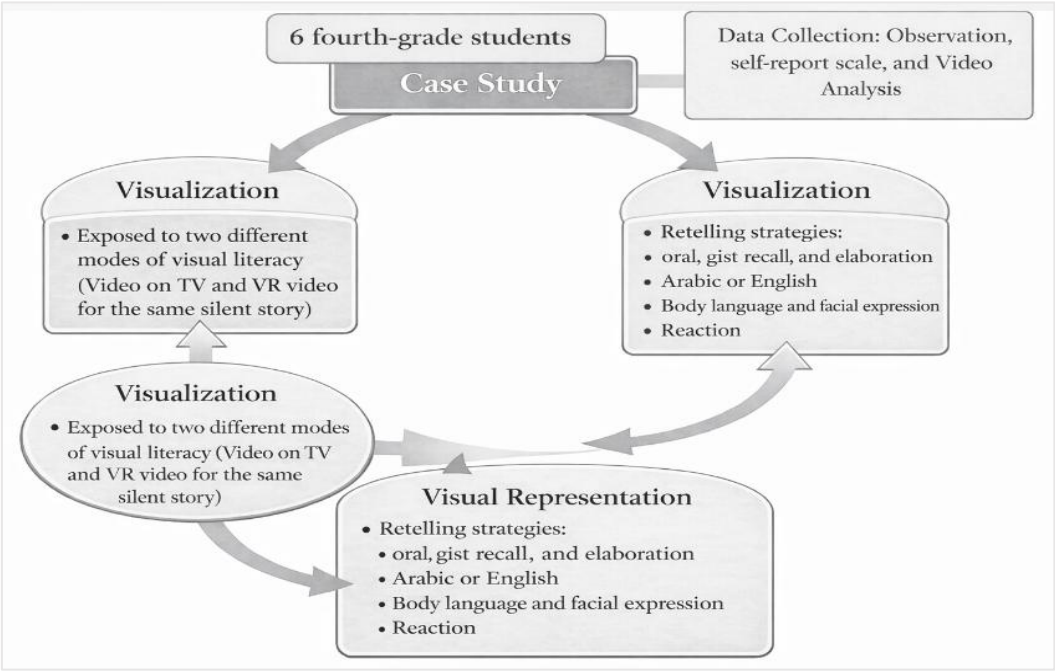
The researchers employed video recordings to capture students' engagement and expressive behaviors while they watched the silent visual story presented in two modalities: TV and VR, and while they were retelling the story. This approach, widely recognized in educational research, enables detailed interpretation of student behavior, particularly engagement and expressive responses. The recording facilitated a comprehensive analysis of verbal and non-verbal cues, including facial expressions, body language, gestures, and posture. Such detailed documentation serves as a robust, objective data source that can be revisited for coding, validation, and further interpretation. This significantly enhances the depth and accuracy of the findings, reinforcing inter-rater reliability and ensuring triangulation across the methodological tools (Derry et al., 2010).

3.4. Procedure

3.4.1. Data Collection

The mixed-methods design enabled the study to examine visual literacy through the complementary components of visualization and visual representation (see Figure 1). This design was theoretically informed by visual literacy theory, which emphasizes learners' ability to interpret and construct meaning from visual texts, and multimodal literacy theory, which views meaning-making as distributed across visual, linguistic, embodied, and affective modes. In the visualization phase, students were introduced to two modes of visual storytelling: TV and VR. The TV mode was selected to represent a conventional visual literacy environment, while the VR mode was included to represent a multimodal literacy context that integrates visual, spatial, and embodied interaction. Both modes presented the same silent narrative to ensure content consistency across conditions.

Figure 1
Methodological Framework



During each viewing session, students' engagement and behavioral responses were systematically observed using an observational checklist. The checklist was designed to capture indicators derived from visual literacy theory (e.g., attention to visual cues, narrative sequencing, and inferential responses) as well as multimodal literacy theory (e.g., embodied reactions, gesture use, and affective engagement). Following each visual presentation, students completed a self-report scale to reflect on their experiences, perceptions, and emotional engagement with the story. This self-report data was

included to capture students' awareness of their visual and multimodal meaning-making processes. In addition, video recordings were made throughout the sessions to support later qualitative analysis and provide a more detailed account of students' interactions with each visual mode. Video data were treated as a primary source for documenting multimodal meaning-making, allowing for the analysis of simultaneous verbal, visual, and embodied actions. This dual-mode procedure allowed for a direct comparison between a conventional visual medium and an immersive digital technology, offering insight into how varying levels of interactivity and immersion influence students' engagement and comprehension during silent storytelling experiences.

During the visual representation phase, students were asked to retell the story orally, with a focus on gist recall and narrative elaboration. Students were free to use either Arabic or English during the retelling process, according to their language preference and comfort. Observational assessments during this phase emphasized both verbal and nonverbal indicators of comprehension, including oral expression, body posture, facial expressions, gestures, and reactions to story events. These indicators were selected to reflect core assumptions of multimodal literacy theory, which recognizes language, gesture, and bodily expression as interrelated meaning-making resources. These observations provided insight into how students constructed meaning and expressed understanding across the two visual modes. This approach is consistent with prior research emphasizing the role of oral and visual storytelling in supporting linguistic and cognitive development, particularly in bilingual learning contexts (Barua, 2023; Ghafar, 2024).

3.4.2. Data Analysis

The study consists of two sequential phases of data analysis. The first phase involved analyzing quantitative data from observational checklists and a self-report scale. In the observation checklist, three researchers rated students' performance on a 1-10 scale, with 1 indicating the lowest and 10 the highest, based on the specific criteria provided in the checklist. These criteria operate key constructs from visual literacy theory (e.g., recall accuracy and narrative coherence) and multimodal literacy theory (e.g., expressive use of gesture and embodied engagement). This evaluation was conducted while the students used both TV and VR modes during their retellings. Descriptive statistics were calculated by tabulating the mean and standard deviation for both TV and VR modes to assess the distribution of each criterion. Given the small sample size and ordinal nature of the 10-point rating scale used in this study, the Wilcoxon signed-rank test was employed to assess whether there were significant differences between the two visual modes in retelling silent stories. Differences between conditions were interpreted in relation to the affordances emphasized by multimodal literacy theory, particularly the role of immersion and embodiment in supporting meaning construction. Although parametric tests like the t-test are known to be robust under certain conditions, non-parametric methods are more suitable when assumptions of normality and interval-level measurement may not be fully met. In this exploratory study, the Wilcoxon test was used to compare paired responses across conditions without making assumptions about the data distribution (Laerd, 2018). For the self-report scale, descriptive statistics were compiled to present students' perspectives on using TV versus VR during silent story retelling. These results were analyzed as reflective indicators of students' visual and multimodal literacy experiences.Their responses were graphed to illustrate the differences between the two modes.

The second phase involved a qualitative analysis of video recordings using thematic analysis, focusing on students' retelling products and linguistic features. The videos were reviewed multiple times to ensure data familiarization, after which meaningful segments were inductively coded, capturing both verbal and nonverbal behaviors, including language use, gestures, facial expressions, body language, elaboration, and reactions (Knoblauch et al., 2014). The coding process was theoretically informed by visual literacy theory and multimodal literacy theory, guiding attention to how meaning was constructed across visual, linguistic, and embodied modes. To enhance analytical rigor, two trained raters independently coded the video data and compared their coding to identify areas of agreement and discrepancy. Through discussion and refinement, codes were organized into broader categories and consolidated into themes representing recurring patterns in students' meaning-making. Within the themtic analysis, sequential analysis was used to trace how communicative actions and narrative construction unfolded over time. This sequential approach aligns with multimodal literacy theory by

examining how meaning-making develops dynamically across time and modes. Reflexivity guided the iterative interpretation process, with each thematic decision continuously reviewed and refined in relation to earlier analyses. Together, sequential analysis and reflexivity ensured that the final themes were systematically derived, collaboratively validated, and firmly grounded in the video data (Knoblauch et al., 2014).

4. Results

4.1. Quantitative Results

To address the first research question, participants' performances were analyzed via an observational checklist as they retold the silent story in two visual modalities (TV and VR). The checklist items were grouped into three categories: expressive language and recall (oral expression and gist recall), attentional and embodied engagement (body posture, focus, and facial expression), and interactive responsiveness (gestures and reactions). This categorization facilitated a systematic comparison between the two modes.

As Figure 2 and Table 2 show, in the TV mode, results indicate generally low mean scores across three analytical categories. Expressive language and recall averaged 4.17, reflecting weak narrative articulation, with elaboration on oral expression at 3.83 and gist recalling at 4.50. Attentional and embodied engagement were also low, with watching attainment at 3.50 and body posture and focus at 2.67, resulting in a category mean of 2.88. Facial expression and body language scores were similarly low. Interactive responsiveness was the weakest category, averaging 2.25, indicating that students engaged passively with the narrative. High standard deviations in some areas suggest variability in individual engagement levels with the TV-based story.

Compared to the TV mode, the VR mode demonstrated consistently higher mean scores across all engagement categories. For expressive language and recall, oral expression averaged 9.17, and gist recalling averaged 8.83, resulting in a mean of 9.00. This indicates that students articulated story details more effectively and retained key elements in VR. Attentional and embodied engagement also improved, with an attainment score of 9.33, and body posture and focus averaged 8.83, while facial expression averaged 9.67. Body language scored 9.00, resulting in a category mean of 9.21, highlighting the immersive nature of VR. Interactive responsiveness was highlighted by an average of 9.00 for gestures and 9.50 for event reactions, resulting in a mean of 9.25. The lower standard deviations in VR measures suggest a more consistent engagement pattern among participants.

Figure 2

Comparison of Category Mean Scores: TV vs. VR

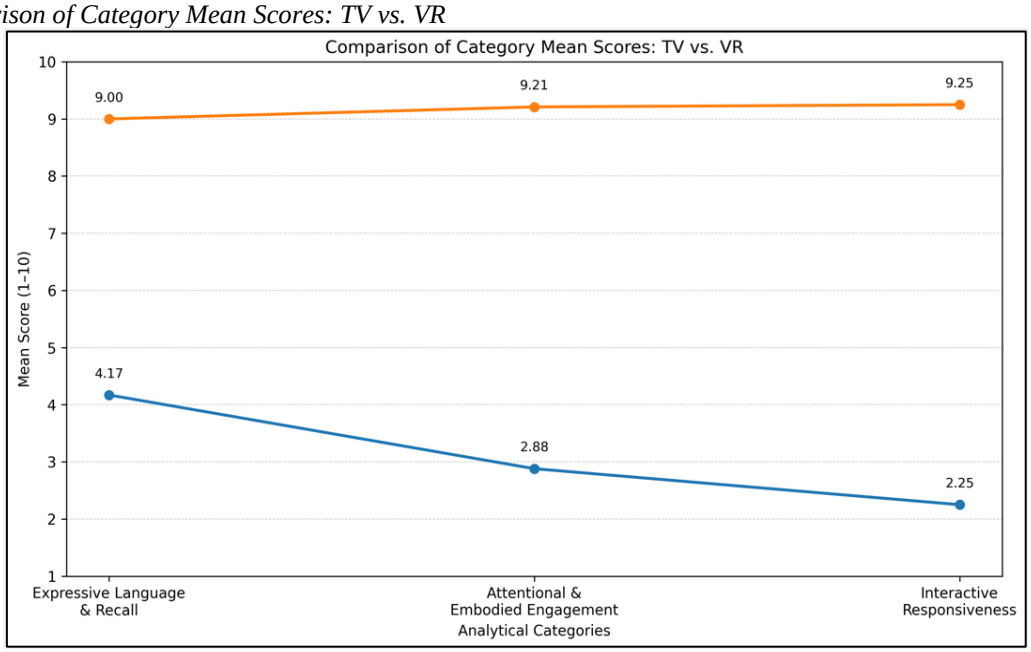


Table 2
Descriptive Statistics: TV vs. VR (for the Sub-Categories)

Main Categories	Sub-Categories	TV		VR	
		Mean	SD	Mean	SD
Expressive Language and Recall	Elaboration on Oral Expression	3.83	2.31	9.17	.98
	Gist Recalling	4.50	1.77	8.83	1.17
Attentional & Embodied Engagement	The Level of Watching Attainment	3.50	1.22	9.33	1.21
	Body Posture & Focus	2.67	1.75	8.83	1.47
	Facial Expression	3.00	.89	9.67	.51
	Body Language	2.33	.81	9.00	1.55
Interactive Responsiveness	Interactive Gestures	2.16	1.16	9.00	1.55
	Reaction to Events	2.33	1.03	9.50	.84

Table 3 presents the results of the related-samples Wilcoxon signed-rank test comparing participants' performance and engagement in the TV and VR conditions. The analysis revealed statistically significant differences between TV and VR across all measured criteria, including expression, gist recall, expressive language, attainment, body posture, facial expression, body language, attentional engagement, interactivity, reaction, and interactive responsiveness. All significance values ranged between .026 and .027, which are below the conventional .05 threshold, indicating that the differences observed between the two conditions are unlikely to have occurred by chance. These findings suggest that the mode of presentation (TV versus VR) had a significant effect on participants' expressive, behavioral, and engagement-related outcomes, although the direction of these differences cannot be determined from Table 3 alone.

Table 3
Related-Samples Wilcoxon Signed Rank Test between TV and VR

Criteria	Sig.
TV –VR Expressive Language and Recall	**.026**
TV – VR Elaboration on Oral Expression	.026
TV – VR gist recall	.026
TV–VR Attentional & Embodied Engagement	**.027**
TV – VR The Level of Watching Attainment	.026
TV – VR Body Posture & Focus	.027
TV – VR Facial Expression	.026
TV– VR body language	.027
TV – VR Interactive Responsiveness	**.026**
TV– VR Interactive Gestures	.027
TV– VR Reaction to Events	.026

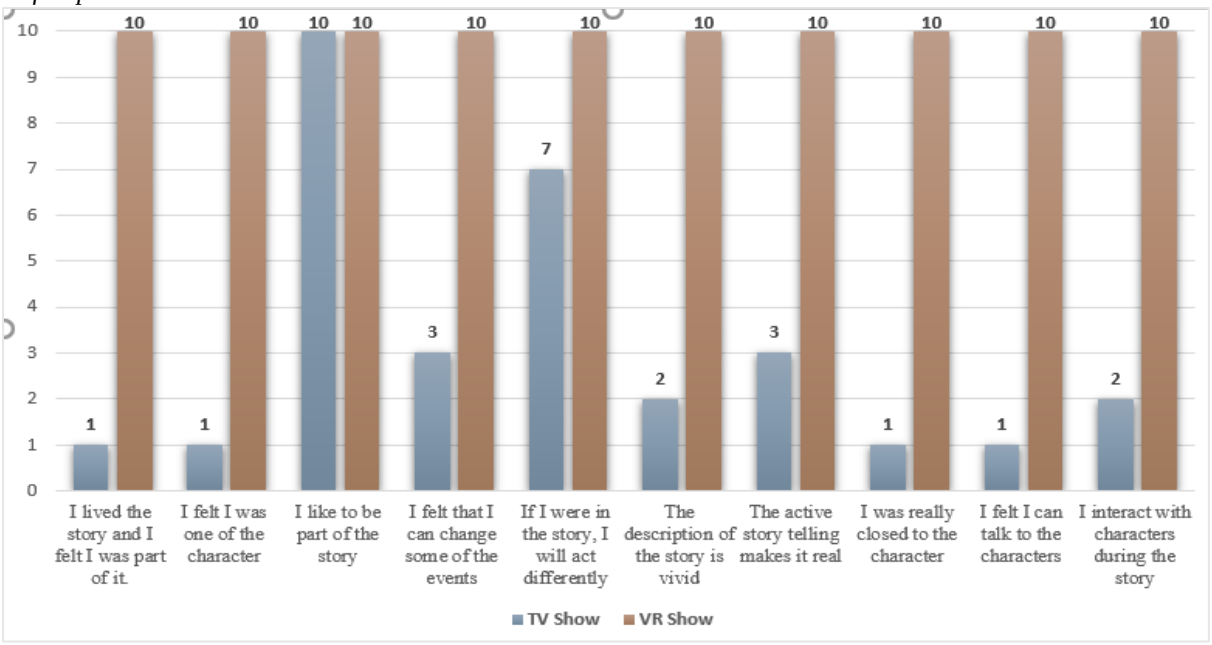
To address the second research question, a self-report scale examined students' subjective storytelling experiences across two visual modes. The items were analyzed across four dimensions: immersion, character identification, perceived agency, and story realism. This approach allowed for a systematic interpretation of students' responses.

In Figure 3, students' responses in the VR mode clustered at the top of the scale across all dimensions. Immersion-related items, such as feeling part of the story, received the highest ratings, indicating that students found VR storytelling engaging. In contrast, the TV mode ratings were significantly lower (1-3), reflecting limited immersion and a passive viewing approach. Students identified strongly with characters in VR and felt they could influence story events, suggesting a sense of narrative agency. In the TV mode, low ratings indicated a fixed narrative with minimal personal involvement.

Students perceived significant differences in story realism and emotional connection between the two modes. The VR mode achieved top ratings for vividness and storytelling realism, indicating a more immediate and engaging narrative. High scores for character closeness suggested that the immersive environment enhanced emotional and social connections. In contrast, the TV mode received low scores

in these areas, indicating that its linear format limited emotional engagement with the story and characters.

Figure 3
Self-Report Scale Results



The analysis reveals that VR's immersive features fostered perceptions of participatory, vivid, and emotionally engaging storytelling among students, while TV elicited more passive responses. These findings highlight how different visual storytelling modes impact emotional, cognitive, and narrative engagement.

4.2. Qualitative Analysis

In response to the third research question, an in-depth video analysis was conducted to examine how VR influenced students' engagement and reading-related behaviors within a bilingual learning context. The analysis was guided by sequential analysis, which enabled the researchers to trace students' actions and responses as they unfolded over time, and by reflexivity, which focused on how students signaled, framed, and contextualized their own participation during the storytelling sessions. This dual analytic approach enabled a nuanced interpretation of both observable behaviors and the meanings students attributed to their actions.

Theme 1: Immersive Environments Support Sustained Engagement and Active Language Use

Sequential video analysis revealed clear differences in the progression of students' engagement across the two visual modes. In the VR condition, students' interactions followed a sustained, continuous sequence of attention, beginning with immediate orientation toward the virtual environment and extending to prolonged visual focus, bodily alignment with the narrative space, and spontaneous verbal and gestural responses. For example, students frequently leaned forward, extended their hands, or shifted their bodies when objects appeared to be close in the virtual scene, often reporting that they felt as if they were "touching" or "holding" something. These embodied reactions occurred immediately following visual cues, indicating a direct, immersive response to the VR environment. In contrast, engagement sequences observed during TV viewing were more fragmented, characterized by limited movement, intermittent attention, and fewer spontaneous reactions, reflecting a less immersive engagement trajectory.

From a reflexive perspective, students in the VR condition consistently framed themselves as participants within the story rather than as external observers. This positioning was evident in their

spontaneous gestures mirroring characters' actions, their verbal reactions to unfolding events, and their self-initiated language use during viewing and retelling. Students frequently spoke aloud to characters or reacted verbally as the narrative progressed, with these behaviors unfolding sequentially across the session rather than appearing as isolated moments. During story retelling, this reflexive positioning led to a greater willingness to elaborate and experiment with language, as students drew on both Arabic and English to express meaning. Together, these patterns suggest that the immersive VR environment supported sustained engagement and active language use by enabling students to experience the narrative as an embodied and participatory event within a bilingual context.

Theme 2: Multisensory Engagement Facilitates Comprehension and Narrative Recall

Sequential analysis of the retelling process revealed clear differences in how students reconstructed the narrative across the two visual modes. Students who experienced the story through VR produced retellings with clearer temporal ordering, stronger internal coherence, and more consistent causal links between events. Their narratives frequently followed the story's original sequence and included detailed references to characters, actions, and settings. In contrast, retellings following the TV condition tended to be shorter and more fragmented, relying on isolated story elements with fewer elaborations and less consistent sequencing.

From a reflexive perspective, VR students signaled their comprehension through a combination of verbal markers, such as emphasis and repetition, and embodied cues, including pointing to imagined locations or reenacting character movements. These gestures often occurred immediately before or during verbal elaboration, indicating that embodied recall supported narrative expression. Language use during retelling further reflected this deeper engagement. Students attempted to retell the story primarily in English, and when lexical gaps or hesitations arose, they strategically shifted to Arabic to clarify meaning or maintain narrative flow before returning to English. This sequential pattern of language shifting reflects translanguaging practices in which Arabic functioned as a supportive resource for meaning making rather than a substitute language. These observations indicate that the multisensory features of VR helped students construct, sustain, and re-experience the narrative during retelling. This support led to deeper comprehension, more coherent storytelling, and a greater recognition of their abilities as bilinguals.

An additional pattern emerging from the video analysis concerns the nature of visual storytelling. The use of a silent story composed solely of visual events appeared to expand students' ways of thinking during retelling. Without relying on verbal narration, students were required to interpret visual cues, infer meaning, and construct the narrative from their individual perceptions of the images. As a result, each student's retelling reflected a unique interpretation of the same visual sequence, demonstrating creative and critical engagement in translating visuals into narrative events. Rather than reproducing a fixed storyline, students drew on imagination, inference, and personal meaning-making to transform visual experiences into spoken narratives. Collectively, these observations indicate that the multisensory features of VR not only supported students in constructing, sustaining, and re-experiencing the narrative during retelling but also fostered deeper comprehension, more coherent storytelling, and greater recognition of their capacities as bilingual meaning-makers.

Theme 3: Emotional Engagement Shapes Participation across Visual Modes

Sequential video analysis revealed distinct trajectories of emotional engagement across the two visual modalities. In the VR condition, students' emotional involvement intensified progressively as the story unfolded, evidenced by increased facial expressiveness, spontaneous reactions to narrative events, and sustained affective responses. These emotional sequences frequently coincided with moments of heightened interaction, suggesting a close relationship between emotional engagement and active participation. In contrast, emotional responses during TV viewing were less frequent and less sustained, with students maintaining a more detached and observational stance throughout the session.

From a reflexive perspective, students in the VR condition consistently framed their engagement through visible affective and linguistic cues that indicated emotional investment in the narrative. Video recordings captured students smiling, laughing, and expressing surprise or concern at key moments,

often vocalizing these reactions spontaneously. During retelling, this emotional involvement was reflected in students' use of evaluative language, such as describing characters as "scared," "happy," or "angry," accompanied by shifts in tone and facial expression. In several instances, students strategically drew on Arabic expressions to convey emotional intensity when English vocabulary was insufficient, before returning to English to continue the narrative. This flexible use of language suggests that emotional engagement prompted translanguaging practices that enabled students to express complex affective meanings without disrupting narrative flow. Notably, some students who initially reported a preference for TV's familiarity demonstrated higher levels of emotional and cognitive engagement during VR sessions, as evidenced by their recorded behaviors. This divergence between stated preference and observed engagement highlights the value of reflexive video analysis in capturing dimensions of participation that may not be fully articulated through self-report alone.

Collectively, the three themes illustrate how VR-supported visual storytelling shaped students' engagement, comprehension, and participation as an interconnected process rather than as isolated outcomes. The immersive environment in Theme 1 sustained students' attention and positioned them as active participants in the narrative, thereby supporting deeper comprehension and more coherent narrative reconstruction as highlighted in Theme 2. This cognitive engagement was further reinforced by heightened emotional involvement, as indicated in Theme 3, in which students' affective responses appeared to energize participation and expressive output. Across all themes, engagement unfolded sequentially and reflexively, demonstrating that students' bodily, linguistic, and emotional responses mutually informed one another. Rather than merely responding to visual input, students actively constructed meaning through interaction, interpretation, and expression, highlighting the role of immersive visual storytelling as a catalyst for integrated cognitive, emotional, and linguistic engagement in a bilingual context. Table 4 summarizes the themes.

Table 4
Themes' Summary

Theme	Analytical Focus	Key Indicators from Video Analysis	Illustrative Evidence
Theme 1: Immersive Environments Support Sustained Engagement and Active Language Use	How immersion shapes attention, bodily engagement, and language use	Sustained visual focus, bodily alignment with the narrative, spontaneous gestures, verbal reactions during viewing	Students leaned forward, extended hands toward virtual objects, mirrored character actions, and spoke aloud to events or characters during VR sessions
Theme 2: Multisensory Engagement Facilitates Comprehension and Narrative Recall	How multisensory input supports narrative reconstruction and meaning making	Coherent temporal sequencing, causal links between events, embodied recall cues, and translanguaging during retelling	Students reenacted story actions while retelling, used gestures to indicate locations, narrated primarily in English, and shifted to Arabic to bridge lexical gaps
Theme 3: Emotional Engagement Shapes Participation Across Visual Modes	How affective involvement influences participation and expression	Facial expressiveness, spontaneous emotional reactions, evaluative language use, shifts in tone and affect	Students laughed, showed surprise or concern during VR, described characters' emotions, and used Arabic expressions to convey emotional intensity before returning to English

5. Discussion

The primary objective of this study was to examine how two visual storytelling modes, VR and TV, support visual literacy, narrative comprehension, and story retelling among fourth-grade bilingual learners in a silent storytelling context. By integrating quantitative measures of retelling performance and engagement with qualitative video-based analysis, the study moves beyond generalized claims about immersion to examine how specific media affordances shape learners' cognitive, emotional, and

linguistic engagement. Grounded in multimodal and visual literacy theories, story retelling was conceptualized as a multimodal outcome reflecting learners' ability to interpret visual cues, organize narrative meaning, and translate visual experience into coherent oral narratives.

The significantly higher VR scores across expressive language and recall, attentional and embodied engagement, and interactive responsiveness strongly align with multimodal literacy theory, which views meaning-making as emerging from the coordinated orchestration of visual, spatial, linguistic, and interactive modes (Kress, 2003; Gee, 2014). Unlike television's fixed viewpoint and linear narrative flow, VR afforded learners agency to navigate narrative space, orient their bodies, and respond dynamically to visual events. This finding supports Kress and van Leeuwen's (2020) argument that control over modal interaction fundamentally shapes how meaning is designed. The qualitative analysis further demonstrated that engagement in VR unfolded continuously and sequentially rather than episodically, suggesting that VR restructures not only the intensity but also the temporal organization of attention. In this sense, VR does not merely enhance engagement; it alters the conditions under which multimodal meaning-making is sustained, supporting embodied cognition and narrative inference in ways that non-immersive screen-based media cannot (Gee, 2014; Merchant et al., 2014; Pilgrim & Pilgrim, 2016).

From a visual literacy perspective, the findings contribute to ongoing debates about how children interpret and produce meaning from images in silent narrative contexts. Visual literacy theory emphasizes that images function as culturally and socially constructed semiotic systems that require inference, emotional interpretation, and contextual reasoning rather than surface recognition (Barthes, 1985; Berger, 2015; Serafini, 2014). The varied yet coherent retellings observed in the VR condition support this view, as learners did not reproduce a fixed storyline but constructed individualized interpretations of shared visual experiences. This aligns with research on wordless and visual narratives, which highlights the role of synthesis, prediction, and narrative coherence in comprehension (Arizpe & Styles, 2015; Pantaleo, 2025). The immersive qualities of VR appeared to intensify these interpretive demands by situating learners within the visual narrative, thereby increasing their responsibility as meaning-makers.

The observed translanguaging practices during retelling further extend the literature by illustrating how bilingual learners flexibly mobilize their linguistic repertoires to mediate visual meaning. Consistent with Garcia and Wei (2014) and recent work on bilingual visual storytelling (Alhabbash, 2022; Alhabbash et al., 2025), translanguaging functioned as a strategic resource rather than a compensatory behavior. VR's rich visual scaffolding appeared to legitimize this flexible language use, supporting narrative continuity, expressive precision, and learner agency. In contrast, the TV condition more often elicited surface-level recall and descriptive retellings, suggesting that reduced visual immersion may constrain opportunities for deeper multimodal and linguistic integration.

Overall, the findings extend research on visual literacy, silent storytelling, and immersive media by empirically demonstrating how VR reshapes the cognitive, linguistic, and semiotic demands of narrative comprehension and retelling. While previous studies have emphasized VR's motivational and affective benefits (Papanastasiou et al., 2019; Wu et al., 2020), this study shows that immersive environments fundamentally reconfigure how learners interpret, organize, and express narrative meaning. Compared with television, which positioned learners as observers, VR positioned them as active designers of meaning, aligning with contemporary calls to reconceptualize literacy as multimodal, embodied, and participatory (Kress, 2003; Serafini, 2024). These findings underscore the importance of examining not only whether immersive technologies engage learners, but also how they transform the processes through which visual narratives are comprehended, retold, and translated into language, particularly in bilingual educational contexts.

The findings of this study have important implications for literacy instruction, particularly in bilingual and visually rich learning environments. First, they suggest that immersive visual storytelling can serve as a powerful pedagogical approach for developing visual literacy and narrative comprehension without relying on written text, making it especially suitable for learners navigating multiple languages. Second, the contrast between VR and TV underscores the need to move beyond assumptions that exposure to visual media alone enhances learning; rather, the nature of interactivity, embodiment, and learner

agency is critical. For educators, integrating VR into literacy instruction offers opportunities to design experiential, student-centered learning tasks that support embodied engagement, emotional investment, and multimodal expression. These findings also highlight the importance of teacher preparation in visual and multimodal literacy, enabling educators to guide students in critically interpreting, discussing, and retelling visual narratives rather than passively consuming them.

While this study provides valuable insights, several directions for future research are warranted. Given the small-scale case study design, future studies should examine larger, more diverse samples to assess the generalizability of these findings across age groups, linguistic backgrounds, and educational contexts. Longitudinal research is also needed to investigate the sustained impact of immersive VR storytelling on reading comprehension, narrative development, and bilingual identity formation. Additionally, future studies could explore alternative interpretations by examining potential challenges associated with VR use, such as cognitive overload or accessibility constraints, and compare VR with other interactive digital storytelling platforms. Finally, further research should investigate instructional scaffolds that support critical visual literacy, ensuring that immersive technologies are used not only to engage learners but also to deepen their analytical and reflective engagement with visual narratives.

Acknowledgments

The authors used an AI-based tool solely for linguistic polishing and/or for generating figures or tables. The ideas, interpretations, and analyses presented in this paper are entirely original and the sole responsibility of the authors.

Disclosure Statement

The authors claim no conflict of interest.

Funding

The research did not receive any specific grants from funding agencies.

References

Al Mohammedi, N. A. (2022). *Learning by design: The transformative pedagogical experience of college students in extending monomodal texts into multimodal* [Doctoral dissertation, United Arab Emirates University]. https://scholarworks.uaeu.ac.ae/all_dissertations/289/

Alhabbash, M. (2022). The noise, the resistance, and the message in construing the construction of Arab bilingual-biliterates' trans-identity space: A grounded theory approach [Doctoral dissertation, United Arab Emirates University]. https://scholarworks.uaeu.ac.ae/all_dissertations/292/

Alhabbash, M., Ibrahim, A., & Al Sheikh, N. (2025). Storying of Arabic bilingual-biliterate identity reconstruction. *Journal of Literacy Research*, *57*(4) 441–470. https://doi.org/10.1177/1086296X251401501

Arizpe, E., & Styles, M. (2015). *Children reading picture books: Interpreting visual texts*. Routledge. https://doi.org/10.4324/9781315683911

Avgerinou, M. D., & Pettersson, R. (2011). Toward a cohesive theory of visual literacy. *Journal of Visual Literacy*, *30*(2), 1-19. https://doi.org/10.1080/23796529.2011.11674687

Barnard, M. (2001). *Approaches to understanding visual culture*. Palgrave Macmillan.

Barthes, R. (1977). *Image, music, text* (S. Heath, Trans.). Hill and Wang.

Barua, S. (2023). Digital storytelling: Impact on learner engagement and language learning outcomes. *International Journal of Academic and Applied Research*, *7*(6), 24-39.

Berger, J. (2015). *About looking*. Bloomsbury Publishing.

Brumberger, E. R. (2011). Visual literacy and the digital native: An examination of the millennial learner. *Journal of Visual Literacy*, *30*(1), 19–47. https://doi.org/10.1080/23796529.2011.11674683

Bryman, A. (2016). *Social research methods*. Oxford University Press.

Christie, F. (1997, October). *Literacy myths*. Education Review.

Creswell, J. W., & Guetterman, T. C. (2024). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research*. Pearson.

Creswell, J. W., & Plano Clark, V. L. (2023). Revisiting mixed methods research designs twenty years later. In C. N. Poth, J. W. Creswell, & V. L. Plano Clark (Eds.), *Handbook of mixed methods research designs* (pp. 21–36). SAGE. https://doi.org/10.4135/9781529682663.n6

Daghriri, I., Liang, C. J., & Huang, W. (2024, May). Spatial immersive learning environments in primary education: A review of impacts and implementation challenges. In T. Huang, L. Hung, & C. Lin (Eds.), *Proceedings of the 2024 International Conference on Information Technology, Data Science, and Optimization* (pp. 99-105). The Association for Computing Machinery. https://doi.org/10.1145/3658549.3658572

Derry, S. J., Pea, R. D., Barron, B., Engle, R. A., Erickson, F., Goldman, R., Hall, R., Koschmann, T., Lemke, J. L., Sherin, M. G., & Sherin, B. L. (2010). Conducting video research in the learning sciences: Guidance on selection, analysis, technology, and ethics. *Journal of the Learning Sciences*, *19*(1), 3–53.

Dooley, K. (2024). Exploring virtual bodies and invisible avatars as storytelling tools in contemporary narrative-based virtual reality projects. *Interactive Film & Media Journal*, *4*(1-2). https://doi.org/10.32920/ifmj.v4i1-2.2061

Eisner, E. W. (1997). Cognition and representation: A way to pursue the American dream? *Phi Delta Kappan*, *78*(5), Article 348.

Ekoç, A. (2021). Teaching speaking with works of art in a preparatory class at university. *Innovation in Language Learning and Teaching*, *15*(5), 415-427. https://doi.org/10.1080/17501229.2020.1824232

Emery, L., Flood, A., Baguley, M., & MacDonald, A. (2019). Visual literacy. In J. Dyson (Ed.), *More than words can say. A view of literacy through the arts* (pp. 18–32). National Advocates for Arts Education.

Farrar, J., Arizpe, E., & Lees, R. (2024). Thinking and learning through images: A review of research related to visual literacy, children's reading and children's literature. *Education 3-13*, *52*(7), 993-1005. https://doi.org/10.1080/03004279.2024.2357892

Gadelha, J. (2023). Imaginar novos mundos agora: Cosmopolíticas e racialidade [Imagining new worlds now: Cosmopolitics and raciality]. *Cadernos de Campo: Revista de Ciências Sociais*, *23*, Article e023021.

García, O., & Wei, L. (2014). *Translanguaging: Language, bilingualism and education.* Palgrave Macmillan.

Gee, J. P. (2014). *Unified discourse analysis: Language, reality, virtual worlds and video games*. Routledge.

Ghafar, Z. (2024). Storytelling as an educational tool to improve language acquisition: A review of the literature. *Journal of Digital Learning and Distance Education*, *2*(10), 781-790. https://doi.org/10.56778/jdlde.v2i9.227

Huang, T. H., Ferraro, F., Mostafazadeh, N., Misra, I., Agrawal, A., Devlin, J., Girshick, R., He, X., Kohli, P., Batra, D., Zitnick, C. L., Parikh, D., Vanderwende, L., Galley, M., & Mitchell, M. (2016, June). Visual storytelling. In K. Knight, A. Nenkova, & O. Rambow (Eds.), *Proceedings of the 2016 conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (pp. 1233-1239). Association for Computational Linguistics.

Ibourk, E. M., & Al-Adwan, A. (2019). Subtitling virtual reality into Arabic: Eye tracking 360-degree video for exploring viewing experience. *Lebende Sprachen*, *64*(2), 286-308. https://doi.org/10.1515/les-2019-0016

Jamaluddin, R. A. D., Hassan, H., Hasan, H., Sayed, D. F. J., Shah, H., & Kamarudin, K. N. (2024). Digital citizenship strategies for the development of visual interactive storytelling applications. *Muhammad Rashaad Bakashmar Noor Inayah Ya'akub*, *10*(2.27), 186-196.

Kendeou, P., McMaster, K. L., Butterfuss, R., Kim, J., Bresina, B., & Wagner, K. (2020). The inferential language comprehension (iLC) framework: Supporting children's comprehension of visual narratives. *Topics in Cognitive Science*, *12*(1), 256-273. https://doi.org/10.1111/tops.12457

Knoblauch, H., Tuma, R., & Schnettler, B. (2014). Video analysis and videography. In U. Flick (Ed.), *The SAGE handbook of qualitative data analysis* (pp. 435-449). SAGE.

Kress, G. (2003). *Literacy in the new media age*. Routledge.

Kress, G., & van Leeuwen, T. (2021). *Reading images: The grammar of visual design* (3rd ed.). Routledge.

Laerd. (2018). *Wilcoxon Signed-Rank Test using SPSS statistics*. https://statistics.laerd.com/spss

Liu, J. (2013). Visual images interpretive strategies in multimodal texts. *Journal of Language Teaching and Research*, *4*(6), 1259-1264. https://doi.org/10.4304/jltr.4.6.1259-1263

Makransky, G., & Petersen, G. B. (2019). Investigating the process of learning with desktop virtual reality: A structural equation modeling approach. *Computers & Education*, *134*, 15-30. https://doi.org/10.1016/j.compedu.2019.02.002

Makransky, G., & Petersen, G. B. (2021). The cognitive affective model of immersive learning (CAMIL): A theoretical research-based model of learning in immersive virtual reality. *Educational Psychology Review*, *33*(3), 937-958. https://doi.org/10.1007/s10648-020-09586-2

Merchant, Z., Goetz, E. T., Cifuentes, L., Keeney-Kennicutt, W., & Davis, T. J. (2014). Effectiveness of virtual reality-based instruction on students' learning outcomes in K-12 and higher education: A meta-analysis. *Computers & Education*, *70*, 29-40. https://doi.org/10.1016/j.compedu.2013.07.033

Metros, S. E. (2008). The educator's role in preparing visually literate learners. *Theory Into Practice*, *47*(2), 102-109. https://doi.org/10.1080/00405840801992264

Mirzoeff, N. (2015). *How to see the world*. Pelican Books.

Morrow, D. G. (1985). Prominent characters and events organize narrative understanding. *Journal of Memory and Language*, *24*(3), 304-319. https://doi.org/10.1016/0749-596X(85)90030-0

(NMC). (2016). *Horizon report: 2016 higher education edition*.

(OECD). (2023). *PISA 2022 results (Volume I): The state of learning and equity in education*. OECD Publishing. https://doi.org/10.1787/53f23881-en

Olshansky, B. (2008). Making magic: Bringing words and pictures together. *Talking Points*, *20*(1), 10–14.

Pantaleo, S. (2025). The meaning-making in kindergarten children's visual narrative compositions. *Journal of Early Childhood Literacy*, *25*(2), 336-366. https://doi.org/10.1177/14687984231161114

Papanastasiou, G., Drigas, A., Skianis, C., Lytras, M., & Papanastasiou, E. (2019). Virtual and augmented reality for the development of spatial and social abilities in children with autism. *Education Sciences*, *9*(2), 1–14.

Paris, A. H., & Paris, S. G. (2003). Assessing narrative comprehension in young children. *Reading Research Quarterly*, *38*(1), 36-76. https://doi.org/10.1598/RRQ.38.1.3

Paris, S. G., & Hamilton, E. E. (2014). The development of children's reading comprehension. In S. E. Israel & G. G. Duffy (Eds.), *Handbook of research on reading comprehension* (pp. 56-77). Routledge. https://doi.org/10.4324/9781315759609

Pauwels, L. (2020). *Visual cultures in science and education*. Springer.

Peixoto, B., Pinto, R., Melo, M., Cabral, L., & Bessa, M. (2021). Immersive virtual reality for foreign language education: A PRISMA systematic review. *IEEE Access*, *9*, 48952-48962. https://doi.org/10.1109/ACCESS.2021.3068858

Pilgrim, J. M., & Pilgrim, J. (2016). The use of virtual reality tools in the reading-language arts classroom. *Texas Journal of Literacy Education*, *4*(2), 90-97.

Pratt, M., & Searles, G. E. (2017). Using visual aids to enhance physician-patient discussions and increase health literacy. *Journal of Cutaneous Medicine and Surgery*, *21*(6), 497-501. https://doi.org/10.1177/1203475417715208

Rowsell, J., McLean, C., & Hamilton, M. (2012). *Visual literacy*. Routledge.

Ruszev, S. (2021). Tracing embodied narrative in VR experiences. *Cinergie–Il Cinema E Le Altre Arti*, *10*(19), 9-20. https://doi.org/10.6092/issn.2280-9481/12334

Serafini, F. (2012). Reading multimodal texts in the 21st century. *Research in the Schools*, *19*(1), 26-32.

Serafini, F. (2014). *Reading the visual: An introduction to teaching multimodal literacy*. Teachers College Press.

Serafini, F. (2024). Developing metamodal awareness. *Children's Literature in English Language Education*, *12*(2), 1-19.

Serafini, F., & Gee, E. (2017). *Visual literacy*. Teachers College Press.

Somadayo, S., & Jamil, M. (2025). Evaluation of online learning system based on virtual reality technology for developing reading literacy in elementary schools. *Mathematical Modelling of Engineering Problems*, *12*(5), 1731-1740. https://doi.org/10.18280/mmep.120528

Sturken, M., & Cartwright, L. (2001). *Practices of looking: An introduction to visual culture*. Oxford University Press.

Sweidan, S. Z., & Darabkh, K. A. (2018). VREG: A virtual reality educational game with Arabic content using android smart phone. *Journal of Software Engineering and Applications*, *11*(10), 500-520. https://doi.org/10.4236/jsea.2018.1110030

Tillmann, A. (2012). What we see and why it matters: How competency in visual literacy can enhance student learning. *Honors Projects*, *9*, 1-29.

Walsh, M. (2021). *Multimodal texts in digital environments: Strategies for teaching and learning*. Routledge.

Wu, B., Yu, X., & Gu, X. (2020). Effectiveness of immersive virtual reality using head-mounted displays on learning performance: A meta-analysis. *British Journal of Educational Technology*, *51*(6), 1991–2005. https://doi.org/10.1111/bjet.13023

Zakime, A. (2017). Using Pinterest to promote genuine communication and enhance personalised learning. In K. Donaghy & D. Xerri (Eds.), *The image* (pp. 51-58). Gutenberg Press.

Zhang, R., Xue, M., & Lysaker, J. T. (2022). Between and within pictures: Gaps and imagination in comprehending wordless books. *Literacies Across the Lifespan Journal*, *2*(1), 3–12.

Appendices

Appendix A

Observational Checklists

Item	**TV Rating 1-10**			**VR Rating 1-10**		
Elaboration on Oral Expression	*Rater1*	*Rater 2*	*Rater 3*	*Rater1*	*Rater 2*	*Rater 3*
Gist Recalling						
The Level of Watching Attainment						
Body Posture & Focus						
Facial Expression						
Body Language						
Interactive Gestures						
Reaction to Events						

Appendix B

Self-Report Scale

Items	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*
1. I felt that I lived the story and was part of it.										
2. I felt that I was one of the characters in the story.										
3. I enjoyed being part of the story.										
4. I felt that I could change some of the events in the story.										
5. If I were in the story, I would act differently.										
6. The description of the story felt vivid and clear.										
7. Active storytelling made the story feel real.										
8. I felt emotionally close to the character(s).										
9. I felt that I could talk to the character(s).										
10. I interacted with the character(s) during the story.										