TURUN YLIOPISTO
UNIVERSITY OF TURKU

Studying illustrated texts and videos in primary classrooms: Learning outcomes, cognitive load, studying processes, and individual differences

Mikko Haavisto

TURUN YLIOPISTON JULKAISUJA – ANNALES UNIVERSITATIS TURKUENSIS
SARJA – SER. B OSA – TOM. 755 | HUMANIORA | TURKU 2025

TURUN YLIOPISTO
UNIVERSITY OF TURKU

STUDYING ILLUSTRATED TEXTS AND VIDEOS IN PRIMARY CLASSROOMS: LEARNING OUTCOMES, COGNITIVE LOAD, STUDYING PROCESSES, AND INDIVIDUAL DIFFERENCES

Mikko Haavisto

TURUN YLIOPISTON JULKAISUJA – ANNALES UNIVERSITATIS TURKUENSIS
SARJA – SER. B OSA – TOM. 755 | HUMANIORA | TURKU 2025

University of Turku

Faculty of Education
Department of Teacher Education
Educational Sciences
Doctoral Programme on Learning, Teaching and Learning Environments Research (OPPI)

Supervised by

Professor Janne Lepola
University of Turku
Turku, Finland

Associate Professor Tomi Jaakkola
Tampere University
Tampere, Finland

Reviewed by

Professor Kristian Kiili
Tampere University
Tampere, Finland

Professor Eliane Segers
Radboud University
Nijmegen, the Netherlands

Opponent

Professor Kristian Kiili
Tampere University
Tampere, Finland

ISBN 978-952-02-0470-9 (PRINT)
ISBN 978-952-02-0471-6 (PDF)
ISSN 0082-6987 (Print)
ISSN 2343-3191 (Online)
Painosalama, Turku, Finland 2025

UNIVERSITY OF TURKU
Faculty of Education
Department of Teacher Education
Educational Sciences
MIKKO HAAVISTO: Studying illustrated texts and videos in primary classrooms: Learning outcomes, cognitive load, studying processes, and individual differences
Doctoral Dissertation, 152 pp.
Doctoral Programme on Learning, Teaching and Learning Environments Research (OPPI)
December 2025

ABSTRACT

Despite the prevalent use of multimedia in primary education, there is limited understanding of how different learning material modalities, such as videos and illustrated texts, affect learning outcomes in classroom settings. Understanding these effects is crucial for educators to tailor instructional strategies that meet the needs of students to facilitate learning. The modality effect is a phenomenon where people learn better from pictures and narration (e.g., videos) than from pictures and written text (e.g., illustrated texts). This dissertation comprises three empirical studies that investigate the modality effect in primary school classrooms. Specifically, it examines how illustrated texts versus videos impact key learning metrics such as retrieval, retention, transfer, and cognitive load, among fifth and sixth graders in Finnish basic education. Additionally, it assesses how studying rate, studying strategies, and reading skills affect these outcomes.

Study I explored the modality effect in a learner-paced classroom setting with a crossover experiment. In the experiment, 54 fifth graders learned science topics from both illustrated texts and videos. They completed pretests, posttests, and delayed tests to assess cognitive load and learning outcomes. The combination of the within-subjects design with linear mixed-effects analysis quantified individual difference factors. The results showed that the video condition outperformed the illustrated text condition in retention, delayed retention, cognitive load, and efficiency measures, although no significant differences were found in transfer tasks. These results show that the modality effect can occur in an authentic primary school classroom environment. Previous studies have often explained that the learning benefits of videos are based on the illustrated narrations causing less cognitive load than illustrated texts. Contrary to this traditional explanation, cognitive load did not emerge as a significant predictor of learning outcomes in Study I, challenging the widely held assumption that videos' learning advantages stem from their ability to reduce cognitive load. This discrepancy points to the need for an alternative explanation for the modality effect.

Motivated by the above findings, Study II aimed to explore factors other than cognitive load for explaining the outperformance of videos. Past research, including Study I, has predominantly focused on learning outcomes without collecting

information on the learning process. Therefore, Study II analyzed 97 screen recordings that were captured on children's iPads during the learning phase of Study I. The analysis examined how children learn from videos and illustrated texts. In particular, studying rates and strategies were assessed; that is, how fast and in what order the children studied the materials. The results demonstrated that videos were studied faster than illustrated texts, and this faster studying predicted improved learning outcomes and lower cognitive load. With both modalities given the same amount of study time, faster studying allowed students to spend more time reviewing the material, providing a plausible explanation for the superior learning outcomes in the video condition. Analysis of the studying strategies provided an explanation for faster studying in the videos: children tended to watch the videos linearly at their designed rate with minimal navigation, and this rate of the videos was faster than the children's reading speed, particularly for weaker readers. Also, the children reviewed mostly by navigating to the start of the video and watching it again from start to finish. While the studying rate in videos was highly homogenous, the studying rates varied widely with illustrated texts. Also review strategies were more varied with illustrated texts. While the videos progressed on their own linearly forward unless interacted with, the illustrated texts required children to actively navigate among slides. This added effort and decision making may explain the higher cognitive load in the illustrated texts.

The slower studying rate in illustrated texts suggests that children's weak reading skills may explain the modality effect, since videos do not require reading. However, Studies I and II lacked direct measurement of reading skills to confirm this.

Study III was designed to address the above gap by investigating the roles of decoding ability and reading comprehension in learning from videos versus illustrated texts. Additionally, Study III aimed to enhance the ecological validity of previous studies by allowing students access to the instructional materials during the posttest, better reflecting typical classroom conditions. A within-subjects crossover experiment involving 109 fifth to sixth graders replicated the results of Study I: videos led to better retention and lower cognitive load. Decoding ability and reading comprehension were both found to be positive predictors of learning outcomes in both modalities. Importantly, an interaction between reading skills and modality was observed: videos provided greater benefits for delayed retention among students with weaker reading abilities. Interestingly, despite prior research suggesting that poor decoding increases cognitive load in text-based materials, there was no significant interaction between reading skills and modality with respect to cognitive load. This suggests that the cognitive load benefits of videos lie elsewhere.

Together, the results challenge the prevailing explanation that the modality effect arises from avoiding cognitive overload. Cognitive load did not explain learning differences between modalities, overall levels of cognitive load were well below capacity limits, and decoding ability did not explain cognitive load differences between modalities. Instead, the results point to alternative explanations and suggest that the modality effect is shaped by higher-level, context-dependent mechanisms. In the classroom context, a key mechanism appears to be the guided and linear nature of studying videos, which standardizes the studying rate and coverage across the content.

The main results of this dissertation have practical value for primary education. Videos can lead to better learning outcomes, lower cognitive load, and faster studying compared to illustrated texts. Moreover, they can promote more equitable learning by inducing more uniform studying, reducing variation in learning outcomes, and supporting students with weaker reading skills. Whereas most previous studies on the modality effect were conducted in highly controlled settings outside the classroom, this dissertation demonstrates these effects in authentic classroom contexts, offering practical information for educators. As producers of learning material often provide educators with both video and illustrated text versions of their content, the findings help teachers make informed decisions in everyday classroom situations.

KEYWORDS: Multimedia learning, video, illustrated text, cognitive load, experimental study, primary education

TURUN YLIOPISTO
Kasvatustieteiden tiedekunta
Opettajankoulutuslaitos
Kasvatustiede
MIKKO HAAVISTO: Kuvitettujen tekstien ja videoiden opiskelu perusopetuksen luokissa: oppimistulokset, kognitiivinen kuormitus, opiskeluprosessit ja yksilölliset erot
Väitöskirja, 152 s.
Oppimisen, opetuksen ja oppimisympäristöjen tutkimuksen tohtoriohjelma (OPPI)
Joulukuu 2025

TIIVISTELMÄ

Multimedian laajasta käytöstä huolimatta on vain rajallisesti tietoa siitä, miten eri oppimateriaalien esitysmuodot, kuten video ja kuvitettu teksti, vaikuttavat oppimistuloksiin luokkatilanteissa. Näiden vaikutusten ymmärtäminen on olennaista, jotta opetusta voidaan räätälöidä oppilaiden tarpeiden mukaan ja edistää oppimista. Modaliteettivaikutuksella tarkoitetaan ilmiötä, jossa ihmiset oppivat paremmin kuvista ja puhutusta tekstistä (esim. video) kuin kuvista ja kirjoitetusta tekstistä (esim. kuvitettu teksti). Tämä väitöskirja koostuu kolmesta empiirisestä tutkimuksesta, jotka tarkastelevat modaliteettivaikutusta alakoulussa. Erityisesti väitöskirjassa tutkitaan, miten kuvitetut tekstit ja videot vaikuttavat keskeisiin oppimismittareihin: tiedonhakuun, muistamiseen, siirtovaikutukseen ja kognitiiviseen kuormitukseen. Lisäksi se arvioi, miten opiskelunopeus, opiskelustrategiat ja lukutaito selittävät näitä tuloksia.

Tutkimus I tarkasteli modaliteettivaikutusta luokkahuonetilanteessa ristikkäisasetelmaa hyödyntäen. Kokeelliseen tutkimukseen osallistui 54 viidennen luokan oppilasta, jotka opiskelivat luonnontieteen sisältöjä sekä kuvitettujen tekstien että videoiden avulla. He suorittivat esitestit, jälkitestit ja viivästetyt testit, joissa arvioitiin kognitiivista kuormitusta ja oppimistuloksia. Toistomittausasetelma ja lineaarisiin sekamalleihin perustuva analyysi mahdollistivat yksilötekijöiden arvioimisen. Tulokset osoittivat, että videot johtivat kuvitettuja tekstejä parempiin tuloksiin muistamisessa, viivästetyssä muistamisessa, kognitiivisessa kuormituksessa ja tehokkuusmittareissa, mutta siirtovaikutuksessa ei havaittu merkitseviä eroja. Havainnot osoittavat, että modaliteettivaikutus voi ilmetä alakoulun luokkaympäristössä. Aiemmassa kirjallisuudessa videoiden hyötyjä on usein selitetty sillä, että kuvien ja puhutun tekstin yhdistelmä aiheuttaa vähemmän kognitiivista kuormitusta kuin kuvien ja kirjoitetun tekstin. Perinteisestä selityksestä poiketen kognitiivinen kuormitus ei kuitenkaan noussut merkitseväksi oppimistulosten ennustajaksi Tutkimuksessa I, mikä haastaa yleisen oletuksen, jonka mukaan videoiden oppimishyödyt johtuvat ensisijaisesti kuormituksen vähentymisestä. Tämä osoittaa, että modaliteettivaikutukselle tarvitaan vaihtoehtoinen selitys.

Edellä mainitut havainnot asettivat tavoitteet Tutkimukselle II, joka tarkasteli muita kuin kognitiiviseen kuormitukseen liittyviä tekijöitä modaliteettivaikutuksen

selittämiseksi. Aiempi tutkimus, myös Tutkimus I, on keskittynyt pääosin oppimistuloksiin keräämättä tietoa itse opiskeluprosessista. Siksi Tutkimuksessa II analysoitiin 97 näyttötallennetta, jotka kerättiin lasten iPad-tableteista Tutkimuksen I opiskeluvaiheen aikana. Analyysissa tarkasteltiin, miten lapset opiskelevat videoista ja kuvitetuista teksteistä. Erityisenä kiinnostuksen kohteena olivat opiskelunopeus ja opiskelustrategiat, eli kuinka nopeasti ja missä järjestyksessä materiaaleissa edettiin. Tulokset osoittivat, että videoiden opiskelu on nopeampaa kuin kuvitettujen tekstien, ja tämä nopeampi eteneminen ennusti parempia oppimistuloksia sekä matalampaa kognitiivista kuormitusta. Kun molempia modaliteetteja oli sama määrä aikaa opiskella, nopeampi eteneminen mahdollisti enemmän aikaa kertaamiselle, mikä saattaa selittää videoiden paremmat tulokset. Opiskelustrategioiden analyysi puolestaan tarjosi selityksen nopeammalle etenemiselle videoissa: lapset katsoivat videot tyypillisesti lineaarisesti niiden alkuperäisellä nopeudella, ja videot etenivät nopeammin kuin lapset lukivat, erityisesti heikompilukutaitoisten kohdalla. Videoita kerrattiin pääasiassa siirtymällä takaisin alkuun ja katsomalla video uudelleen alusta loppuun. Videoissa etenemisnopeus oli hyvin homogeeninen, kun taas kuvitettujen tekstien opiskelunopeudet vaihtelivat suuresti. Myös kertaamisstrategiat olivat kuvitetuissa teksteissä monimuotoisempia. Videot etenevät omatoimisesti lineaarisesti eteenpäin, ellei käyttäjä puutu toistoon, mutta kuvitettujen tekstien opiskelu edellyttää aktiivista siirtymistä sivujen välillä. Tämä ylimääräinen aktiivisuus ja päätöksenteko saattavat osaksi selittää kuvitettujen tekstien aiheuttaman korkeamman kognitiivisen kuormituksen.

Kuvitettujen tekstien hitaampi eteneminen viittaa siihen, että lasten heikot lukutaidot voisivat olla selitys modaliteettivaikutukselle, koska videoiden katsominen ei edellytä lukutaitoa. Tutkimuksissa I ja II ei kuitenkaan mitattu lukutaitoa, jotta tämä oletus voitaisiin todentaa.

Siksi Tutkimus III selvitti dekoodaustaidon ja luetun ymmärtämisen yhteyttä videoiden ja kuvitettujen tekstien opiskeluun. Lisäksi Tutkimus III pyrki vahvistamaan aiempien tutkimusten ekologista validiteettia sallimalla opiskelijoille pääsyn opetusmateriaaliin jälkitestin aikana, mikä vastaa paremmin tyypillistä luokkahuonetilannetta. Toistomittausasetelmaan perustuva ristikkäiskoe, johon osallistui 109 viidennen ja kuudennen luokan oppilasta, toisti Tutkimuksen I tulokset: videot johtivat parempaan muistamiseen ja matalampaan kognitiiviseen kuormitukseen. Sekä dekoodaustaito että luetun ymmärtäminen ennustivat positiivisesti oppimistuloksia molemmissa modaliteeteissa. Lukutaidon ja modaliteetin välillä havaittiin vuorovaikutus: videot olivat erityisen hyödyllisiä viivästetyn muistamisen kannalta niille oppilaille, joiden lukutaito oli heikompi. Vaikka aiempi tutkimus kertoo heikon dekoodaustaidon lisäävän kirjoitetun materiaalin kognitiivista kuormitusta, lukutaidon ja modaliteetin välillä ei havaittu merkittävää vuorovaikutusta kognitiivisen kuormituksen osalta. Tämä viittaa siihen, että videoiden vähempi kuormittavuus johtuu muista tekijöistä.

Kokonaisuudessaan tulokset haastavat vallitsevan selityksen, jonka mukaan modaliteettivaikutus perustuu kognitiivisen ylikuormituksen välttämiseen. Kognitiivinen kuormitus ei selittänyt oppimistulosten eroja modaliteettien välillä, kokonaiskuormitus pysyi selvästi kapasiteettirajojen alapuolella, ja dekoodaustaito ei selittänyt kuormituseroja modaliteettien välillä. Sen sijaan tulokset viittaavat vaihtoehtoisiin selityksiin ja tukevat tulkintaa, että modaliteettivaikutus muotoutuu korkeamman tason kontekstisidonnaisten mekanismien kautta. Luokkahuonetilanteessa keskeinen mekanismi vaikuttaa olevan videoiden ohjaava ja lineaarinen katselutapa, joka yhtenäistää opiskelunopeuden ja tukee sisällön tasapuolista opiskelua.

Väitöskirjan päätuloksilla on käytännön merkitystä perusopetukselle. Videot voivat parantaa oppimistuloksia, vähentää kognitiivista kuormitusta ja nopeuttaa opiskelua verrattuna kuvitettuihin teksteihin. Lisäksi videot voivat edistää oppimisen tasa-arvoa ohjaamalla opiskelua yhtenäisemmäksi, vähentämällä oppimistulosten vaihtelua ja tukemalla erityisesti heikommin lukevia oppilaita. Toisin kuin useimmat aiemmat modaliteettivaikutusta koskevat tutkimukset, jotka on toteutettu kontrolloiduissa laboratorio-olosuhteissa, tämä väitöskirja tutkii vaikutuksia aidossa luokkaympäristössä ja tarjoaa opettajille käytännön tietoa. Koska oppimateriaalien tuottajat tarjoavat usein sekä video- että kuvitettuja tekstiversioita sisällöistään, tulokset auttavat opettajia tekemään perusteltuja valintoja arjen opetustilanteissa.

ASIASANAT: Multimediaoppiminen, video, kuvitettu teksti, kognitiivinen kuormitus, kokeellinen tutkimus, perusopetus

Acknowledgements

There were many people without whom this would not have been possible and to whom I want to express my gratitude. I value each person for the unique role they had in my journey.

I am grateful to my supervisors, Professor Janne Lepola and Associate Professor Tomi Jaakkola, for the trust they placed in my skills. They gave me the freedom to follow the direction that made sense to me, yet did not hesitate to challenge weak points or warn me when the path ahead looked risky. Their willingness to treat me as an equal discussion partner and coauthor shaped how I think and work. Janne showed how things function within the academic world and looked out for my interests at every stage. Tomi provided methodological guidance and set an example of academic argumentation and writing.

I wish to thank the pre-examiners and the opponent of this dissertation. I am honored that they devoted their time and expertise to my work. Professor Kristian Kiili provided extensive feedback that pushed me to revise the text to bring forward the overarching point of the dissertation. Professor Eliane Segers' comments helped me strike a balanced tone in my arguments, and she gave me a solid foundation for developing my methods by sharing materials and tests used in prior research.

I am thankful for the funding from the Doctoral Programme on Learning, Teaching and Learning Environments Research (OPPI). The steady support allowed me to pursue my own vision without interruption and to grow at a natural pace. I appreciate the programme's confidence in me and the opportunities it offered, including support for travel and conferences. I also want to thank the Department of Teacher Education for providing a place where I could work independently and feel part of a supportive academic environment.

I am thankful to everyone who shaped my journey. At the beginning, Anni told me during our runs what a PhD involved. Kati supported me throughout, arranging school connections and developing learning materials that were vital for this dissertation. Xiaoshan was a valuable gym and discussion partner who helped me think through the complexities of education while keeping in shape. I am also grateful to Nina, Lauri, Tiia, Sanni, and Arttu for collaborations that kept my thinking sharp and helped me learn from outside my narrow field. I hope these collaborations continue for a long time. Lastly, I want to thank all colleagues who tolerated me using the coffee room as my office and who shared humor and peer support when it was most needed.

I owe much to my family. My mother, Ani, has been the most reasonable, fair, and selfless person I know, and her example has been invaluable. My brother Timo explained to me how the universe works when I was young, and the aim to understand the world from first principles has stayed with me. Lastly, my partner Bahar has supported me through challenges and shown me how to be courageous in the face of uncertainty and not to let opportunities pass by.

Turku, 13 November 2025
Mikko Haavisto

Table of Contents

Tables

Figures

List of Original Publications

This dissertation is based on the following original publications, which are referred to in the text by their Roman numerals:

I Haavisto, M., Jaakkola, T., & Lepola, J. (2023). Video outperforms illustrated text: Do old explanations for the modality effect apply in a learner-paced fifth-grade classroom context? *Computers & Education*, 199, 104775. https://doi.org/10.1016/j.compedu.2023.104775

II Haavisto, M., & Jaakkola, T. (2025). Fifth graders study videos faster than illustrated texts: Studying rate, review strategies, and learning implications. *PsyArXiv*. https://doi.org/10.31234/osf.io/mztpw_v1

III Haavisto, M., Lepola, J., & Jaakkola, T. (2025). The "simple" view of learning from illustrated texts and videos. *Learning and Instruction,* 100, 102200. https://doi.org/10.1016/j.learninstruc.2025.102200

Haavisto developed the original study ideas and contributed to the conceptualization and experimental design. Haavisto was also responsible for data collection, data analysis, interpretation of the results, and primary manuscript writing. Jaakkola and Lepola acted as supervisors, contributing to the conceptualization and methodology of the study and providing guidance throughout the research process. They also reviewed and edited the manuscripts.

The original publications have been reproduced with the permission of the copyright holders. Study II is under review in *Instructional Science*.

Declaration of generative AI and AI-assisted technologies. During the writing process of this dissertation, Haavisto used ChatGPT to improve language and readability. After using this tool, Haavisto reviewed and edited the content as needed and takes full responsibility for the content of the publication.

1 Introduction

Videos and illustrated texts are among the most common multimedia materials used in schools. Today, most teachers use videos from YouTube or other digital platforms (Hyvönen et al., 2018; Pattier, 2021) and textbook producers provide video content in addition to illustrated texts. However, it is unclear whether these videos are beneficial to primary school children's learning compared to illustrated texts.

Numerous studies have shown that the question of which material is better for learning depends on the context of learning (Ginns, 2005; Reinwein, 2012). That is, who the learners are, what the environment is, and what the materials are like. Consequently, general research-based suggestions like "use spoken rather than written text with pictures" (Moreno & Mayer, 1999) may have limited applicability (Scheiter et al., 2014; Schüler et al., 2012), particularly if the experimental settings differ substantially from real-world classroom practices (Kingstone et al., 2008). Most prior multimedia research has been conducted in laboratory settings with adult participants, using materials that did not allow learner control (i.e., learners could not navigate the material at their own pace). It is therefore unknown to what extent the results generalize to primary school classrooms that use modern, learner-controlled materials.

This dissertation investigates how fifth and sixth graders learn from illustrated texts and videos in the classroom. The illustrated texts combined pictures with written text, while the videos used the same pictures and words, with the words presented as narration instead of written text. Both materials were accessed on tablets and allowed learner control. The goal was to understand how these materials differ in supporting learning. The research examines how factors such as individual differences, studying behaviors, reading skills, and material characteristics explain differences in learning outcomes and cognitive load. The theoretical aim was to test existing explanations for the benefits of narrated and written multimedia in realistic classroom environments and to propose alternative explanations.

Laboratory research typically aims to control all variables. However, such control can introduce systematic bias when the goal is to generalize results to uncontrolled real-world settings. This dissertation takes a different approach by accounting for natural variation through study design and statistical methods. First, both illustrated texts and videos were studied at the same time in the same classroom, which controlled for known and unknown environmental factors. Second, because every child learned from both videos and illustrated texts, the design accounted for known and unknown individual differences. This approach aims to preserve the authentic classroom context while maintaining scientific rigor.

Together, the theoretical aims and methodological approach allow the dissertation to provide practical information that generalizes to classroom environments. The practical aim was to provide evidence that supports teachers in choosing between videos and illustrated texts and understanding the reasons behind that choice.

2 Theoretical Background and Aims

2.1 Multimedia learning and the modality effect

The modality effect refers to improved learning when multimedia materials (i.e., materials with words and pictures) present information through two sensory modalities rather than just one (Low & Sweller, 2014). As videos utilize both vision for pictures and audition for narration, the modality effect posits that they would support learning better than illustrated texts, which use only vision for both pictures and text. The modality effect has been studied extensively, and empirical syntheses have confirmed the learning benefits of narrated multimedia (Ginns, 2005; Mayer, 2020; Reinwein, 2012). This evidence is mostly based on experiments that compare learning outcomes after learners have studied using either written or narrated multimedia.

Reinwein (2012) distinguishes between micro- and macro-level modality effects, but these phenomena are often discussed interchangeably in the literature, leading to confusion. Modality effect research started at the micro-level with studies that used very brief verbal items and tightly controlled laboratory tasks. The separate-streams hypothesis (Penney, 1989) and the dual-coding theory (Paivio, 1969) were often used as explanations for the modality effect. Many of the experiments relied on materials that only lasted a few seconds (Low & Sweller, 2014). Therefore, it could be argued that a more accurate term for this effect would be the auditory recency effect (Rummer et al., 2011), as it might be based on short-term advantages of auditory sensory memory (Section 2.2). However, this might be less relevant for the macro-level modality effect concerning long and complex materials (Low & Sweller, 2014). This dissertation focuses on the macro-level modality effect because the micro-level effect is likely overshadowed by other higher-level effects in classroom learning (Schüler et al., 2011). In addition, the dissertation highlights how the micro-level origins of modality effect research continue to influence contemporary macro-level studies. This mismatch between research paradigms makes it unclear to what degree the results of modality effect research are valid, especially when generalizing them to authentic learning situations.

Most often, the modality difference is defined as the contrast between written and narrated text, with all other factors held constant. However, there are multiple interpretations of what "modality" refers to in the context of the modality effect. That is, what aspects of the materials are changed when comparing different modalities. This makes comparisons between studies challenging. Some studies have altered the presentation technology, delivery medium, and even learner control as part of the modality difference. For example, in Herrlinger (2017), narration was presented through a CD-player for the whole class at a given pace, while the written text was on paper, freely browsable by the students. In this dissertation, such differences in delivery medium and learner control are not seen as part of the modality difference: children studied both videos and illustrated texts on iPads, free to browse both materials.

However, videos and illustrated texts do differ in some fundamental ways, which should be taken into account when aiming for a fair comparison between the modalities. Written text can be navigated simply by shifting one's gaze across the text, whereas different parts of the narration cannot be heard by directing one's ears. This creates an inherent difference between the modalities, leading to different navigation options being implemented in illustrated texts and videos. Videos have a specific timeline that can be rewound or fast-forwarded, and they progress automatically, which is unnecessary in illustrated texts that learners can read at their own pace. Illustrated texts are navigated by switching pages or slides within which the learner uses their eyes for finer pacing control.

As there is disagreement in the field about which features are considered inherent parts of modality, this dissertation adopts generalizability to classroom contexts as its guiding principle. This serves two aims: (a) to provide practical information to teachers and learners choosing between different modalities, and (b) to investigate the theoretical processes underlying the modality effect in an authentic learning environment. To achieve these goals, the illustrated texts and videos were designed to be as similar as possible (e.g., having identical pictures and words), except for the qualities necessary for them to reflect the typical materials used in primary school classrooms (e.g., videos progress automatically but illustrated texts do not).

2.2 Model of memory and learning

Multimedia learning theories such as cognitive load theory and cognitive theory of multimedia learning (Section 2.3) are based on a three-component model of memory. This model, formalized by Atkinson and Shiffrin (1968), includes sensory memory, working memory, and long-term memory.

Information from the senses, such as sight and hearing, first enters sensory memory. Sensory memory is modality-specific and short-lived: visual traces last less than one second (Sperling, 1960), and auditory traces last less than four seconds (Darwin et al., 1972). Because many of the experiments on the micro-level modality effect used materials lasting only a few seconds, the effect was more dependent on sensory memory (Low & Sweller, 2014). In contrast, working and long-term memory play a more central role in the macro-level modality effect.

Information from sensory memory that receives attention is further processed in working memory. Working memory includes a visuo-spatial component and a phonological component (Baddeley, 1992, 2012). Written and narrated words are represented in the phonological working memory, whereas pictures are processed in the visuo-spatial component. Working memory has a limited capacity of about four chunks when chunking and rehearsal are controlled for (Cowan, 2001). Chunking refers to the processing of several familiar or related elements as a single piece of information in working memory (a chunk), which allows more information to be handled within its limited capacity. Main information processing occurs in the working memory, and the fundamental cause of cognitive load is the interaction of multiple chunks in working memory (i.e., element interactivity, Sweller, 2010). Without rehearsal, working memory retains information for less than 30 seconds (Peterson & Peterson, 1959).

Finally, some information processed in working memory is encoded into long-term memory, which can store information for a lifetime (e.g., Bahrick, 1984) and has no practical capacity limit (e.g., Brady et al., 2008). In this dissertation, the retention tests required participants to encode the studied information into long-term memory. To illustrate this, consider a posttest comprising questions about the studied material. If a student relies solely on working memory, they must repeat a specific fact in their minds until they encounter the corresponding question. This strategy would yield only one correct answer, which highlights the role of long-term memory even in relatively short-term tasks. In addition to retention, this dissertation measured retrieval with tests where the participants answered questions by retrieving information from materials. In these tasks, the participants could keep the questions and answers in working memory one by one, reducing demands on long-term memory. Finally, because long-term memory decays over time, a delayed retention test administered one week later was used to assess the durability of the encoded information.

2.3 Cognitive load theory

A dominant theoretical framework for explaining multimedia learning is cognitive load theory (Leahy & Sweller, 2016). A basic explanation of learning effectiveness in cognitive load theory is that processing irrelevant information (i.e., extraneous cognitive load) can be avoided through better instructional practices, which frees working memory to process relevant information (i.e., intrinsic cognitive load), leading to improved learning (Paas & Sweller, 2014). In contrast, learning suffers when the total cognitive load exceeds the learner's cognitive capacity (i.e., cognitive overload, Ginns, 2005). In general, this dissertation aimed to follow common instructional practices to reduce extraneous cognitive load in both modalities, to the extent they are followed in materials used in schools today. For example, the contiguity principle was followed by presenting relevant text and pictures as close to each other as possible spatially and temporally, the coherence principle was followed by excluding words and pictures that were not relevant to the learning goal, and the redundancy principle was followed by excluding subtitles from the videos (for more on multimedia design principles, see Mayer & Fiorella, 2014).

This dissertation focuses on total cognitive load, which is defined as the sum of intrinsic cognitive load (i.e., cognitive load caused by the information that serves learning purposes) and extraneous cognitive load (i.e., cognitive load caused by information that is unnecessary for learning; see Kalyuga, 2011). This two-component model of cognitive load was chosen because it is theoretically difficult to distinguish between intrinsic load and the common third type of cognitive load called germane load. To address this, Kalyuga (2011) and Sweller (2010) proposed that germane load can be seen as part of intrinsic load, resulting in the simpler two-factor model. Jiang and Kalyuga (2020) tested this issue empirically with confirmatory factor analysis of undergraduates' cognitive load during three learning tasks. They used a modified version of Leppink et al.'s (2013) questionnaire, which was developed to measure intrinsic, extraneous, and germane load. However, only two distinct factors, intrinsic and extraneous, were found. They conclude that the dual model of cognitive load offers a more conceptually clear framework. It is also worth noting that early research on cognitive load theory typically referred only to intrinsic and extraneous load and the concept of germane load was introduced later (Kalyuga, 2011).

Cognitive load can be assessed from the perspective of the learning material or the learner. The material perspective assesses cognitive load based on how many elements of information the materials present or how many elements must be processed simultaneously to answer a question (i.e., element interactivity). However, for long and complex materials, it is often impossible to determine the cognitive load potential of the material. This also varies between individuals, as the extent to which learners can chunk information in working memory is not feasible to determine in complex tasks. For these reasons, this dissertation examines cognitive load from the learner's perspective (see Kalyuga, 2007, 2011), understanding it as the mental effort experienced by the learner when they actively process information.

Cognitive load measurement can be categorized into four types: direct objective, indirect objective, direct subjective, and indirect subjective (Jiang & Kalyuga, 2020). Direct objective measures assess cognitive load through observable indicators that reflect working memory activity during task performance (e.g., dual-task performance, eye tracking, fMRI). Indirect objective measures are also observable but assess outcomes or behaviors assumed to result from cognitive load (e.g., EEG, test scores). Direct subjective measures rely on learners' self-assessments of the effort or difficulty experienced during the task (e.g., mental effort ratings). Indirect subjective measures use self-reports of experiences believed to be influenced by cognitive load (e.g., stress ratings). This dissertation chose to measure cognitive load with a direct subjective measure, particularly a single-item rating scale of mental effort (Paas, 1992). The following justifies this choice.

A wide range of modern cognitive load measurement methods and their combinations are presented in Chen et al. (2016). They report cognitive load during four tasks of varying difficulty using subjective ratings, dual-task response time, pupil diameter, galvanic skin response, and blink rate. Multiple figures (Chen et al., 2016; Figures 4.2, 4.3, 4.4, 5.4) compare different measurement approaches. They show that subjective ratings differentiate task difficulty levels consistently, with higher difficulty corresponding to higher ratings. This supports the assumption that self-assessments targeting average cognitive load may provide a useful proxy for overload during learning tasks, even if they cannot reflect rapid load fluctuations. In contrast, measures such as response time, galvanic skin response, blink rate, and pupil diameter generally fail to distinguish between difficulty levels, producing largely overlapping results.

In addition to being highly reliable at distinguishing levels of task demand, subjective ratings are easy to implement in authentic classroom settings, and they do not interfere with the main learning tasks (Jiang and Kalyuga, 2020). The Paas (1992) single-item scale also has strong face validity, as it asks learners to rate their amount of mental effort they invested, aligning directly with the construct of cognitive load.

In contrast, many other scales could be argued to have less direct alignment with the construct of cognitive load, as they measure task difficulty (e.g., Marcus, Cooper, & Sweller, 1996, experiment 3; Kalyuga, Chandler, & Sweller, 1998). This may be based on the two different perspectives on cognitive load: (a) learner's experience (mental effort) and (b) the element interactivity of the learning material (task difficulty). Multi-item scales seem to differ even more from the core cognitive load construct. For example, in the Leppink et al.'s (2013) scale, items intended to measure germane cognitive load ask about perceived understanding, extraneous cognitive load items ask about the clarity of the instruction, and intrinsic cognitive load items ask about the complexity of the material. However, none of these items ask learners to assess how effortful the learning experience was or how much cognitive load they experienced, which raises concerns about whether the scale measures cognitive load itself. Lastly, it is questionable whether learners can be expected to detect the subtle differences in phrasing used between questionnaire items. This is even more so for fifth and sixth graders in this dissertation.

While this dissertation focuses on cognitive load theory, the cognitive theory of multimedia learning is another dominant framework for explaining multimedia learning. The two theories overlap substantially and share core assumptions about the structure and limits of working memory, including dual-channel processing and limited capacity. They also account for similar multimedia learning effects, such as the modality effect and the split-attention effect (Section 2.4). To describe cognitive demands, cognitive theory of multimedia learning refers to extraneous, essential, and generative processing instead of extraneous, intrinsic, and germane load; these constructs are analogous (Mayer, 2021). For consistency, this dissertation adopts the terminology of cognitive load theory. The relationship and overlap between the two theories are described in Mayer (2021), and in Castro-Alonso and Sweller (2021) on the modality effect.

2.4 Explanations for the modality effect

The modality effect cannot be attributed to a single cause. Cognitive load theory details multiple distinct mechanisms for why learning from narrated or written multimedia would lead to better learning outcomes or reduced cognitive load. The final observed modality effect reflects the combined influence of positive and negative effects associated with each modality.

Cognitive load theory suggests that presenting information through both visual and auditory channels of working memory reduces cognitive load and therefore enhances learning. There are two ways in which this directly manifests when learners study illustrated texts and videos: decoding and split attention.

Decoding forms the basis for one of the main explanations behind the modality effect. According to cognitive load theory, illustrated narrations may impose less cognitive load during learning than illustrated texts because humans are evolutionarily adapted to process spoken language (Paas & Sweller, 2012). In contrast, written text is read by decoding the text in working memory, which increases cognitive load. This process also requires the retrieval of letter and word patterns from long-term memory, and the learning of these patterns requires extensive systematic practice (Paas & Sweller, 2012). This dissertation focuses on the role of reading skills in the modality effect by including the assessment of decoding and reading comprehension in an experiment where children study videos and illustrated texts. The decoding advantage of narrated multimedia may be especially important for children because their reading skills have not yet fully developed (Ehri, 2005; Reinwein & Tassé, 2022).

Another explanation for the modality effect is based on **split attention**. Videos allow learners to view pictures and listen to narration at the same time, whereas illustrated texts require them to divide their attention between pictures and text (split attention effect; Chandler & Sweller, 1992). From the perspective of working memory, videos allow learners to receive both auditory and visual information simultaneously to verbal and visuo-spatial working memory (Baddeley, 1992; Penney, 1989), whereas in the case of the illustrated text, the visual channel is overloaded with two competing inputs. Therefore, cognitive load is reduced when learning with narrated multimedia, which could produce the modality effect (Tabbers et al., 2004; cf., Moreno & Mayer, 1999; Reinwein & Tassé, 2022).

Split attention in illustrated texts can be reduced up to a point by bringing the words and pictures as close to each other as possible (contiguity principle, Mayer & Fiorella, 2014). For example, in the case of diagrams, the words can often be integrated into the pictures by using labels. However, in contrast to common topics in the modality effect research that focused on learning from diagrams and shorter texts (e.g., Moreno & Mayer, 1999), this approach is not feasible for longer expository texts such as those used in this dissertation. Thus, the occurrence of the split-attention effect depends on the length of the text. Such moderating factors are called boundary conditions; that is, conditions that define when a specific learning effect is expected to appear or disappear. Another boundary condition for the split-attention effect is the relevance of pictures. If the material includes irrelevant pictures and the learner recognizes this, they can focus on just the written text, negating the split-attention effect. In this dissertation, split attention is expected to influence the modality effect because the texts were relatively long, and the pictures were designed to be relevant by including questions in posttests about content found only in the pictures.

While cognitive load theory traditionally predicts a modality effect, several characteristics of written multimedia may benefit learning compared to narrated multimedia, leading to a reverse modality effect. The **transient information** effect provides one of the most discussed explanations for a reverse modality effect. Spoken content is fleeting, requiring learners to continuously process and integrate information in working memory. In contrast, written texts are permanent, allowing rereading and reducing memory demands (i.e., the transient information effect; Singh et al., 2012). Therefore, learning from narrations can be impaired when new information replaces prior information faster than the learner's cognitive capacity allows (Leahy & Sweller, 2011). Furthermore, videos often include dynamic visualizations, which further increase the transient information effect (Merkt et al., 2011). However, the videos and illustrated texts used in this dissertation were matched in picture content and limited to static illustrations. This decision was based on the facts that (a) dynamic and static illustrations cannot be matched in content, and (b) dynamic visualizations are not commonly used in illustrated texts because they cannot be synchronized with permanent written text.

In texts, the learner can (a) effortlessly direct gaze to different parts of the written text and (b) process multiple words simultaneously (Snell & Grainger, 2019), which is much faster and more fluent than navigating video. Furthermore, eye-tracking studies (Biedert et al., 2012; Simola et al., 2008) show that when people locate information, they often engage in a faster type of reading characterized by searching for specific words without attending to meaning (Brysbaert, 2019). This makes locating relevant information easier. In contrast, the learners only hear one instance of the narration at a time. This creates challenges for accessing information fluently in videos. Low and Sweller (2014) explain that learner control can reduce the transience of narrations. Indeed, how useful video navigation is depends on the fluency and flexibility of the available navigation options, which in previous studies have often been limited (Section 2.7).

For some mechanisms it is unclear whether they would benefit learning based on videos (modality effect) or illustrated texts (reverse modality effect). One such mechanism is the **shift of attention**. A systematic review of eye-tracking research suggests that participants learning from spoken multimedia consistently devoted more visual attention to the accompanying images than those who studied written multimedia (Alemdag & Cagiltay, 2018). The impact of such shift of attention to the modality effect again depends on how relevant pictures are. With relevant pictures, videos would benefit, whereas with irrelevant pictures, illustrated texts would benefit. It is plausible that children would shift their attention even more towards pictures because their reading skills are still developing (Roy-Charland et al., 2007).

Finally, it is important to note that the above explanations represent just the most frequently discussed ones in prior research on the modality effect. There are also less discussed explanations, many of which will be discussed in the following sections.

2.5 Reading ability and the modality effect

Over the past decade, children's reading skills have declined markedly on a global scale. According to recent PISA data, 15-year-olds today perform at the reading level of 14-year-olds from ten years ago (OECD, 2023, p. 158). This trend is concerning, as education continues to depend heavily on textbooks. One possible response to this challenge is the use of video-based instructional materials, which may support learners with weaker reading skills. Although reading skills have been suggested as a factor influencing the modality effect, its role has rarely been directly assessed. Prior studies have provided indirect evidence; for example, Knoop-van Campen et al. (2018, 2019) reported stronger modality effects in children with dyslexia than typically developing children. However, those studies did not assess how decoding ability affects learning across modalities. The present study extends prior work by directly examining how children's decoding ability and reading comprehension affect learning from both videos and illustrated texts.

Reading ability is framed using the simple view of reading, which proposes that reading comprehension results from the combined contribution of decoding and linguistic comprehension (Gough & Tunmer, 1986). Decoding involves fluent reading of individual words, while linguistic comprehension refers to understanding sentence meaning based on word-level information. Both elements are essential; neither alone is enough for comprehension. A learner who can decode words but lacks linguistic comprehension may read aloud accurately without grasping the meaning. On the other hand, without decoding skills, the reader lacks access to the word-level information necessary for interpreting the text.

Especially decoding skills play a central role in explaining the modality effect, as cognitive load theory proposes that converting visual symbols into verbal representations imposes additional load on working memory (Wong et al., 2012). Therefore, low decoding skill increases the mental effort needed to process written content.

The separation of decoding in the simple view of reading offers a practical framework for examining how reading skills influence the studying of illustrated texts and videos. Its straightforward structure makes it more practical for classroom use than more complex models. For example, the integrated model of text and picture comprehension (Schnotz, 2014) describes multiple separate processes for transforming written input into phonological representations, while the reading systems network (Perfetti & Stafura, 2014) presents decoding as part of a dynamic and recursive system of interrelated processes.

Better decoding ability often predicts better learning also from spoken materials, even though they do not directly require decoding. This is because decoding and linguistic comprehension are typically positively correlated (e.g., Language and Reading Research Consortium, 2015). Reading and listening comprehension are also correlated, largely because both rely on linguistic comprehension (Torppa et al., 2016). In addition, reading and listening comprehension involve both overlapping and modality-specific processes, and fluent word reading supports comprehension in both modalities (Wolf et al., 2019).

Unlike decoding, linguistic comprehension directly supports learning from both videos and illustrated texts. Therefore, broader measures like reading comprehension or verbal intelligence can benefit both modalities equally. For example, Segers et al. (2008) found that verbal intelligence did not predict the modality effect. Previous research has also associated reading comprehension with improved multimedia learning in general. Herrlinger et al. (2017) found that reading comprehension among fourth graders accounted for a substantial proportion of the variance in retention and transfer outcomes (R^2 = .28–.37). Likewise, Witteman and Segers (2010) reported that sixth graders with stronger reading comprehension achieved higher overall learning outcomes across multimedia formats. However, neither study investigated whether reading comprehension influenced learning differently depending on the modality, leaving unexplored a key interaction addressed in the current research.

Learning to read requires considerable effort and sustained practice, especially for children (Sweller, 2020). For children with weak reading skills, even individual graphemes may represent separate chunks in working memory, easily overloading their cognitive capacity. In contrast, videos bypass the need for decoding and may be particularly beneficial for learners with weaker decoding skills. As decoding becomes automatic, more cognitive resources become available for comprehension (Perfetti, 1985), reducing the cognitive load benefit of listening. Although decoding is essential during the early stages of reading development, its importance decreases as proficiency improves (Smith et al., 2021).

Developmental evidence supports this shift. For example, Reinwein and Tassé (2022) found that adults processed written multimedia messages more efficiently than spoken ones, whereas the reverse was true for sixth graders. Similarly, Holopainen's (2003) longitudinal study showed that Finnish children performed better in listening comprehension than reading comprehension in third grade, but the difference had disappeared by ninth grade. Therefore, listening is particularly effective for learning at least throughout primary school, benefiting video-based learning.

Despite these differences, young learners are severely underrepresented in modality effect research. Only one of the 51 studies included in Reinwein's (2012) meta-analysis examined young learners in a learner-paced setting. In addition, most cognitive load studies have focused on adults (Mutlu-Bayraktar et al., 2019), and research on the modality effect has largely relied on university students (Mayer, 2020), which is a highly selective group with relatively strong reading skills. In contrast, the present research was conducted in public primary schools, providing a more representative sample in terms of reading ability.

2.6 Studying rates in various learning materials

Much of the research on reading rates has centered on university students, who typically read non-multimedia, nonfiction texts silently at an average pace of 248 words per minute (wpm; Brysbaert, 2019). In contrast, children read considerably more slowly. According to Spichtig et al. (2016), sixth-grade students average about 165 wpm. However, these figures largely reflect research conditions where learners are highly focused on a shorter task. In more realistic educational settings, such as during classroom instruction, reading speeds are often reduced due to distractions and lapses in attention (Brysbaert, 2019).

Reading speed is also shaped by the purpose of reading. For instance, Dirix et al. (2020) found that adults who expected to answer retention questions read at just 54 wpm, whereas those unaware of such questions read the same text at 189 wpm. These slower reading speeds for retention can be explained by strategies such as frequent rereading to organize information (i.e., elaborative rehearsal, Craik & Lockhart, 1972). This deeper processing tends to produce more detailed and durable memory representations (Meyer et al., 1999). There are also age differences. Hyönä (1994) estimated that adults read at 106 wpm, compared to 82 wpm for fifth graders, both expecting retention questions.

Although children read at slower rates than adults, they are often able to understand spoken language at comparable speeds (Woodcock & Clark, 1968). This difference stems from the fact that reading requires decoding, which is a skill that develops after language comprehension. As a result, listening may serve as a more effective channel for supporting comprehension in younger learners, especially for those who struggle with decoding.

In contrast, the rate at which adults process optimally sped-up narrations is roughly equivalent to their reading speed (Kuperman et al., 2021). This equivalence aligns with Baddeley's (2012) model of working memory, which suggests that verbal inputs, no matter if they are auditory or visual, are encoded in a shared phonological working memory, placing similar capacity limits on both modalities.

Selecting an appropriate speed for presenting learning materials is not straightforward. Narrations that are too fast may overwhelm working memory, impeding learning (Mayer & Moreno, 2003; cf. Breznitz & Berman, 2003). On the other hand, if narration is too slow, it can hinder comprehension by disrupting the segmentation of information into coherent chunks (Kintsch & Van Dijk, 1978) and by limiting the time available for review activities during lessons.

In multimedia contexts, measuring presentation rate solely in words per minute (wpm) fails to reflect the full scope of delivered information, as it neglects the role of images. Nevertheless, many earlier studies have reported rates in terms of wpm alone (e.g., Pastore, 2010, 2012; ten Hove & van der Meij, 2015; Guo et al., 2014). In this dissertation, **studying rate** refers to the number of new words read or listened to and new pictures seen per minute (wpm + ppm), calculated as an average over the first study-through of the material. The focus is exclusively on new content, as rereading the same slide multiple times does not indicate rapid progression through the material.

The materials children consume today have much more picture content than decades ago. Still, the role of pictures in multimedia learning remains largely neglected compared to the text. Previous research on studying rates has primarily been based on non-multimedia materials (Pastore, 2010). Furthermore, the few available studies involving multimedia are outdated (Brysbaert, 2019) and have focused exclusively on reading, listening, or narration rates, while overlooking the studying rate of pictures (Pastore, 2012). The studies of reading rates of primary school children often do not mention the influence of pictures to reading rates (e.g., Tortorelli, 2019; Taylor, 1965; Spichtig et al., 2016; Brysbaert, 2020), even though informational texts for this age group are commonly full of pictures.

When pictures are added to written texts, they may slow down children's reading because of split attention. In contrast, videos allow learners to process spoken narration and visuals at the same time through separate sensory channels. This implies that videos could be presented at rates comparable to those of audio-only narrations (e.g., Pastore, 2012; Ritzhaupt et al., 2008). However, the presence of pictures increases the total amount of information to be processed. This might also affect video-based learning, leading to a lower studying rate.

Research on the impact of pictures on reading speed has produced mixed findings. For example, Hannus and Hyönä (1999; experiment 2) used eye-tracking to show that fourth-grade students spent just 6% of their study time looking at illustrations. In contrast, Mason et al. (2013) reported that 86% of fourth graders in their eye-tracking study attempted to integrate information from both text and the accompanying image, indicating a much greater focus on illustrations.

The effect of different types of pictures on studying rate is not straightforward either. Photographs, for example, can support learners in quickly understanding the context of the text, but they also include extra details that may distract from the main learning objectives (Rey, 2012; Chang & Choi, 2014). While photographs are commonly used in primary school textbooks, multimedia learning research has tended to focus on diagrams (e.g., break and pump systems, Mayer & Gallini, 1990). Diagrams typically contain dense information and require learners to manage several elements in working memory at once (element interactivity, Sweller, 2010), which can raise cognitive demands and potentially slow down the studying rate.

2.7 Learner control in multimedia learning

In recent decades, the use of personal tablet computers in primary education has made it possible for children to engage with multimedia materials at their own pace. However, there has been limited research on how primary school students navigate videos and illustrated texts, and what learning advantages result from this interaction. Many previous studies have provided learners with either no learner control (i.e., system-pacing) or only limited learner controls, in contrast to the flexible navigation options available in modern educational materials (e.g., timeline slider in videos and navigating forwards or backwards to any slide in illustrated texts; Domagk et al., 2010; Fyfield et al., 2022; Höffler & Schwartz, 2011; Kay, 2012; Reinwein, 2012). However, system-pacing is increasingly uncommon in the classroom, as almost all educational content offers learner control options.

Experiments that compare children's learning performance in narrations and illustrated texts have focused on learning outcomes, while giving little attention to the learning process itself, which limits how the results can be interpreted (Höffler & Schwartz, 2011). The connection between learning processes and outcomes remains insufficiently understood, especially in video-based learning, where prior research is scarce (Lee & List, 2019; List & Ballenger, 2019). Nevertheless, variables such as navigation behavior and time spent on learning are often used as explanations, even though their effect on learning differences between modalities is not directly investigated (e.g., Knoop-van Campen et al., 2018, 2019; Segers et al., 2008; Witteman & Segers, 2010). This dissertation aims to investigate the relationship between process measures and learning outcomes in illustrated texts and videos, thereby advancing the current understanding of how learner control impacts multimedia learning.

The role of learner control also interacts with the transience of the materials. Spoken text is usually given as an example of transient information and contrasted with written text, which is permanent (Ayres & Sweller, 2014). However, the transience of these instructional messages greatly depends on the learner controls. Meta-analyses (Ginns, 2005; Reinwein, 2012) and recent studies in primary school settings (e.g., Knoop-van Campen et al., 2018, 2019) have shown that the availability of learner controls reduces the modality effect, which is unexpected as they mitigate the effects of transience, benefiting narrations. There are multiple possible mechanisms that could lead to this result.

First is the type of learner controls implemented in previous studies. In several studies involving fifth and sixth graders, narration has been controlled on a slide-by-slide basis rather than through continuous playback (e.g., Knoop-van Campen et al., 2018, 2019; Segers et al., 2008; Witteman & Segers, 2010). This design can cause disruptions, as learners must notice when the narration for one slide finishes and then manually advance to and activate the next audio segment. This may shift attention away from studying toward managing the interface, increasing extraneous cognitive load (Hatsidimitris & Kalyuga, 2013). Supporting this, an eye-tracking study by Zhu et al. (2020) showed that learners in narrated conditions spent more time focusing on navigation controls, likely because they were uncertain when the current narration clip would end. In contrast, videos today provide flexible learner controls in addition to fluid automatic progression, which is referred to as hybrid learner control. The narrated condition in this dissertation used hybrid learner control, implemented as YouTube videos.

Another explanation for why learner controls reduce the modality effect is that more learning time and the opportunity to take breaks and review content reduces cognitive overload, therefore making the cognitive load reduction of narrated multimedia unnecessary. For example, Merkt et al. (2022) found that the ability to pause a video allows learners to manage cognitive load during challenging segments. However, they also note that the option to pause may increase cognitive load because the learners are thinking whether or not to pause. It is well known that active navigation is associated with higher cognitive load (Hatsidimitris & Kalyuga, 2013; Scheiter & Gerjets, 2007).

Learner controls may also affect experts' and novices' learning differently (Spanjers et al., 2011). Hatsidimitris and Kalyuga (2013) found that the addition of a timeline slider benefited low prior-knowledge learners with a large effect size, while high-knowledge learners performed slightly better without the slider. They explain that this result is because low prior knowledge learners are specifically vulnerable to transient information because they lack the schemas that reduce processing demands for experienced learners. In contrast, learner controls may be detrimental to novices because they may overlook important content or disrupt the logical flow of explanations by making poor navigational choices (Corbalan et al., 2009; Scheiter, 2014). However, learners with limited prior knowledge often avoid using navigation features because they find it difficult to recognize the structure of the information presented (Amadieu et al., 2009; Biard et al., 2018; Chen et al., 2006; Hasler et al., 2007).

A counterintuitive finding is that adding pacing controls can make video-based learning more active, even if learners do not use those controls. Hasler et al. (2007) found that simply having the option to pause or stop the video promoted active engagement with the content, possibly because it encourages learners to monitor its structure (see also Merkt et al., 2018; Rey et al., 2019; Spanjers et al., 2012). For example, learners may monitor the content for logical breakpoints where pausing would make sense, possibly improving comprehension (Merkt et al., 2022). In this dissertation, the participants represent novices, as the studied topics were specifically chosen to be novel to them.

Active processing is seen to be a core component of multimedia learning (Mayer, 2020). One explanation for why watching instructional videos is generally associated with lower mental effort (Beege et al., 2022) is that reading requires learners to engage more actively with the material, as they must actively move their gaze, whereas listening is more passive because the narration progresses automatically (e.g. Herrlinger et al., 2017; Segers, 2008). A contrary point of view is that listening requires sustained attention and inhibitory control (Kim & Phillips, 2014). In either case, both these aspects, in addition to prior knowledge, make individual differences a key factor for the modality effect. This dissertation analyzes individual differences to investigate their role.

Finally, an issue in many studies of the modality effect is that learner-control has been a confounding factor (see Merkt et al., 2011). For example, studies of fourth to sixth grade students by Leahy & Sweller (2011) and Wong et al. (2012) presented the materials in system-paced slides, and in Herrlinger et al. (2017), the written condition was learner-paced (sheets of paper with pictures and text) while the narrated condition was system-paced (pictures on paper but narration played to the class). Moreover, the first experiment in Leahy et al. (2003) gave learners unlimited learning time in the written condition, whereas the narration could only be listened to once. In all these examples, the written text group could reread the text according to their reading ability, whereas the spoken text group could not relisten. This asymmetry likely favored the text condition and does not reflect modern multimedia applications, where narration can be paused or replayed.

2.8 Studying strategies

In this dissertation, studying strategy is defined narrowly as the way a learner progresses through the material; specifically, the order in which content is accessed and at which rate. Access to learner control enables the use of studying strategies. Furthermore, there is a direct connection between the use of navigation options (e.g., pausing, fast-forwarding, or skipping slides) and the rate and order of studying. Default progression corresponding to passive strategies differs between modalities. When studying videos, the studying rate and order follow the default linear progression of the video, unless navigation options are used. When studying illustrated texts, progressing beyond the first page requires active navigation.

Previous studies have shown that learners are capable of flexibly shifting between reading strategies when engaging with written texts (Goldman & Saul, 1990). They examined three main strategies: frequently revisiting earlier parts of the text, reading through the text once in a linear fashion, and reading the entire text followed by selective rereading. Participants often changed their approach when reading texts, depending on the context.

Similarly, de Boer et al. (2011) found that learners also adapt their strategies when watching videos. They identified three common strategies: watching the video straight through without interruption, rewatching the entire video, and repeatedly viewing specific segments. Encouraging students to reflect on their viewing strategies led to a 20 percent improvement in learning outcomes, and students' awareness of the type of posttest questions also influenced their strategy choices. Michel et al. (2007) found that watching a video twice improved children's retention. They proposed that this benefit may come from becoming more familiar with the video's format and content during the first viewing, which allows for more effective integrative processing during the second.

Slow reading may reflect different underlying strategies, some of which are more effective than others. Hyönä et al. (2002) identified four distinct reading behavior patterns in an eye-tracking study and examined how these patterns predict summarization performance. The highest performing group was selective reviewers, who read at 139 words per minute and often revisited introductory or summary sections of the texts. The lowest performing group was slow linear readers, who read at 133 words per minute. While slow linear reading may hinder comprehension, rereading key sections can be beneficial. Supporting this, studies by Hyönä (1994, 1995) show that adults tend to apply integrative strategies when studying challenging expository texts, especially during the initial reading.

In this dissertation, the learning time was divided into initial pass (i.e., the first study through the material) and review (i.e., the continued study of the same material after the initial pass) phases, and studying strategies were examined separately for both. This was motivated by several prior studies that have identified clear differences in the strategies learners employ between initial pass and review. Kim et al. (2014) found that students tend to watch videos in a more linear fashion during the initial viewing, but during review, they use navigation options to focus on specific segments. Mason et al. (2013) further note that the initial pass reflects simpler processing, while the review phase is characterized by more intentional and strategic engagement. In contrast, Hyönä and Niemi (1990) found that during the initial pass of a text, adults exhibited longer fixation times and more regressions compared to repeated readings. This indicates that more intensive processing occurred during the initial pass, whereas the purpose of review was perhaps more confirmatory.

2.9 Efficiency based on learning, time, and effort

In multimedia research, the modality effect can refer to improvement in various learning-related outcomes. First, it can be observed in learning outcomes, such as when narrated multimedia leads to better retention or transfer compared to written multimedia. Second, the effect has been demonstrated in cognitive load, when the narrated multimedia leads to less mental effort. Third, the modality effect can arise in learning time, when studying the same content takes less time in the narrated condition.

In this dissertation, learning is considered more **effective** when it produces better learning outcomes, requires less time, or is less effortful. Learning is more **efficient** when it produces better learning outcomes relative to the effort and time invested. These three benefits can often be exchanged for one another by varying instructional methods. For example, a learning method that takes less time can be extended to enhance learning outcomes without increasing cognitive load. Similarly, introducing breaks may lengthen learning time but reduce cognitive load, while keeping learning outcomes similar. Conversely, cognitive load can often be exchanged for better learning outcomes by demanding more effort from learners, keeping the learning time constant.

An issue arises when one of these three types of outcomes is evaluated in isolation without accounting for the others, which is common in previous studies. A good example is Knoop-van Campen (2019), where the typically developing group showed nonsignificant reversed modality effects in learning outcomes but a positive modality effect in learning time. This raises the question of which learning material performs better when both are considered together, as gains in time could be offset by disadvantages in learning outcomes. Furthermore, as cognitive load was not measured, it is possible that the shorter learning time might have come at the cost of increased cognitive load, which could tire learners over longer periods.

An illustrative example of the importance of efficiency for both practical applications and theoretical clarity is the comparison between learner-controlled and system-controlled instruction. Learning time commonly varies under learner-controlled conditions, while it is predetermined in system-controlled conditions. Introducing learner control frequently increases the time spent learning, and it is argued to be the main reason for its learning benefits (Spanjers et al., 2010). For instance, Gerjets et al. (2009) reported that allowing learners to control the pace of instruction resulted in a 6% improvement in posttest performance, although this came with a 31% increase in learning time. However, it could be argued that this represents a worse outcome, because the time saved in the system-controlled condition could have been allocated to further studying in a real-world learning context.

While established efficiency calculations exist for combining learning and cognitive load benefits (relative condition efficiency, Paas & Van Merriënboer, 1993) and learning and time benefits (learning efficiency; Rasch & Schnotz, 2009), there is no agreed-upon method for combining learning, cognitive load, and time into a single metric. As learning time is the only one of these that can be practically experimentally controlled, all participants in this dissertation are given a limited time to study the learning materials. Learning outcomes and cognitive load are measured, so the approach accounts for all three factors and facilitates evaluating the real-world benefits of illustrated texts and videos.

In this dissertation, **efficiency metrics** refer specifically to adapted instructional efficiency (Van Gog & Paas, 2008), an adaptation of relative condition efficiency (Paas & Van Merriënboer, 1993). The metric is adapted because it combines test performance with cognitive load during learning rather than during testing (see calculations in Section 4.5). The adapted version was chosen because, in Paas and Van Merriënboer (1993), the testing condition corresponded to one of the learning conditions. In contrast, in the present research, testing and learning phases are entirely distinct, and cognitive load during learning is of primary interest.

A final complication in analyzing learning time is that the time actually used by learners often differs from the provided time in learner-controlled conditions. Prior research has yielded mixed results on how much time learners use when learning time is unrestricted. For example, Tabbers and de Koeijer (2010) reported that learners spent significantly more time studying compared to the duration of the narration, with some students spending more than three times longer, indicating extensive reviewing. In contrast, Leahy, Chandler, & Sweller (2003, exp. 2) reported that sixth graders used less than half of the time in the written condition compared to the length of the narration, even though participants were instructed to read until they were confident that they understood the material. Thus, time used seems to depend on the instruction given to the learners, as well as learner characteristics such as motivation and metacognition.

In this dissertation, the available time for learning was limited; however, after participants initially studied the material from start to finish, they were explicitly encouraged to use all the remaining time for reviewing, mirroring typical classroom instruction.

2.10 Ecological validity

The following example illustrates how the learning process can differ between experimental and classroom conditions.

Imagine a ten-year-old child sitting at their desk in the classroom. A stranger enters and says it is their turn. The child is escorted away from classmates and brought into a quiet room where they sit alone with an unfamiliar adult. They are instructed to listen to a narration while looking at pictures on a computer. The adult presses play, and the narration begins.

The slideshow seems unusual: half of each slide is blank, as if something is missing. Just as the child begins to focus on the task, the narration is paused, and they are asked to assess how difficult the task is. The narration continues, and the child tries to quickly remember what was said before the interruption.

The child feels increasingly aware of their gaze, having been told that a camera is tracking their eye movements. Suddenly, they realize they have been focusing on the camera rather than the narration. Regrettably, there is no way to rewind. For the rest of the narration, the child listens carefully and examines each picture, doing what they think the researcher expects. When the slideshow ends, they answer test questions quickly, motivated by the thought of returning to the comfort of the classroom.

Now consider a typical classroom situation. The child is studying the presentation on their own tablet, sitting at their own desk. Occasionally, they pause the presentation to whisper a funny comment to a classmate and are reminded by the teacher to let everyone work in peace. They might get lost in thought and glance out the window, wondering when recess will begin. They return to the presentation, rewind the part they missed, and continue watching. At some point, they become focused on the content and forget where they are, only realizing it when the presentation ends. This classroom situation is what the present dissertation aims to assess.

Children may perform differently in lab and classroom conditions. Yet much of what we know about how children learn from multimedia materials comes from lab settings (Butcher, 2014). This creates a need for more research involving children in classroom settings (Herrlinger et al., 2017; Mayer & Fiorella, 2014; Schüler et al., 2012; Yang et al., 2013). Lab studies are valuable for isolating specific mechanisms under controlled conditions. However, studies that aim to provide instructional guidelines should be conducted in authentic contexts to show that the results generalize to those contexts (see Kingstone et al., 2008).

While the underlying mechanism for the modality effect and the situations in which it occurs are still debated, studies have certainly demonstrated that the modality effect is context-sensitive (Sections 2.1–2.9). The following sections (2.10.1–2.10.3) outline specific contextual factors that may affect the modality effect. Since there is a lack of prior research comparing learning in research context versus authentic contexts, it is unknown to what degree the context alters the results. However, to generalize results to classrooms, the burden of proof lies on laboratory research to show that none of these factors matters for the modality effect.

One well-known aspect of ecological validity is the temporal aspect, which is widely discussed in research on the novelty effect. It is a "cognitive, affective and/or behavioral change that occurs as a result of initial exposure of study participants to a new situation, such as an experiment or a new learning technology. This change eventually wanes or disappears as participants acclimate to the new situation. Thus, the functioning of individuals during their initial exposure to an experiment should not be used to determine their expected functioning at a later stage or over time" (Klein, 2022, p. 1599). However, other aspects of ecological validity have received less attention in research.

This dissertation aimed to test the modality effect in a setting that reflected the primary school classroom context as a whole with respect to its setting, materials, and tests (i.e., dimensions of ecological validity, Schmuckler, 2001).

2.10.1 Ecological validity of the settings

Apart from lab environments, previous studies on the modality effect in children have taken place in diverse settings, such as one-on-one sessions outside the classroom with a researcher (e.g., Knoop-van Campen et al., 2018, 2019), small group sessions conducted in school hallways or auditoriums (e.g., Segers et al., 2008; Witteman & Segers, 2010), or by assigning different modality conditions to separate classrooms (e.g., Herrlinger et al., 2017). These variations in setting may affect both learning outcomes and cognitive load in multiple ways (Choi et al., 2014).

In the classroom with peers versus one-on-one with a researcher. A within-subjects experimental study showed that children performed better in individual settings in both visuospatial and verbal working memory with a medium effect size compared to classroom settings **(**Friso-van den Bos & van de Weijer-Bergsma, 2020). Working memory performance has direct consequences for cognitive load, and thus the modality effect. This suggests that narrations may be especially beneficial in a classroom setting because they reduce cognitive load.

Children's performance can be influenced by the presence of a researcher during individual sessions. In such contexts, they may experience performance pressure and therefore invest a consistently high level of effort across conditions, regardless of their interest in the material. In contrast, motivation may vary more in less supervised classroom settings (see Eysenck & Calvo, 1992). If, for example, videos are more motivating or require less effort, such benefits could be less apparent in an individual testing setting.

Attention and distraction. Children may direct more attention to an unfamiliar environment, which can increase extraneous cognitive load due to the additional mental effort required to process novel surroundings (Cycowicz, 2019). Additionally, an eye-tracking study showed that in authentic learning settings, students' attention often drifts away from the instructional material (Yang et al., 2013). The study found that the students looked somewhere other than the presentation for 40% to 70% of the viewing time. This suggests that classroom behavior could be less focused and more fragmented than in lab studies.

Physical aspects. Classrooms are purposefully designed to support learning and differ from other environments in physical aspects such as lighting, ambient noise, and the number of people present, all of which can influence learning outcomes (Evans, 2006; Higgins et al., 2005). Particularly the background noise level provides an intuitive example of how physical aspects could affect modality conditions differently. As noise directly impairs listening comprehension (Klatte et al., 2010), a modality experiment in a quiet lab environment could be biased in favor of narrations (see also Vettori et al., 2022).

Familiarity of settings. Much of children's science learning likely occurs in their own classroom, making their prior knowledge context-specific (see Smith &

Vela, 2001). This prior knowledge can for example reduce the modality effect by increasing the redundancy of the information presented (Kalyuga, 2014).

Learning time. In primary classrooms, individual learning often occurs between group activities or the start or end of the lesson, which limits the learning time given to students. In contrast, studies with learner control have often allowed unlimited learning time (Section 2.7), which does not reflect classroom practice. If one modality benefits more from extended study time, this could bias results in favor of that modality.

Another aspect that deviates from classroom practice is that in many system-paced studies, the time given to study illustrated texts is matched to the duration of the corresponding narration (Tabbers et al., 2004; Schüler et al., 2012). This gives an optimal amount of time for the narrated condition, while readers may get too much or too little time depending on their reading speed. Tabbers et al. (2004) confirmed this, showing that a modality effect appeared when text time was matched to 1x narration, disappeared at 2x, and reversed with unlimited time.

2.10.2 Ecological validity of the materials

There are multiple ways of implementing written and narrated multimedia and the differences between these implementations can give an advantage to either condition.

Learner experience is perhaps the most concrete example of how low ecological validity can jeopardize generalization. For example, fifth graders have likely never studied from an illustrated text that progresses at its own pace, or from a video that does not. In contrast, they have five years of systematic practice using ecologically valid types of illustrated texts and videos. Therefore, an observed effect may reflect familiarity with the material format, not differences in modality.

Half-blank screens. In many previous studies of the modality effect, the narrated condition had half of the screen blank where the written text was removed (e.g., Knoop-van Campen et al., 2018, 2019; Segers et al., 2008). This indicates that the illustrated text served as the main reference for the learning material, whereas the narrated version prioritized scientific control rather than ecological validity. Presenting content on just part of the screen removes an inherent benefit of narrated multimedia: it has more space for images than illustrated texts. This may reduce the modality effect. Additionally, this may increase extraneous cognitive load. For example, children might think there is a technical issue with the presentation, or they might exert effort unlearning their habit of scanning the entire screen for information.

Unfamiliar learner controls. The choice of learner controls in illustrated narrations also suggests that the illustrated text has been the primary basis of learning material design. In many studies, the narration was implemented as separate clips in a slideshow presentation. As discussed in Section 2.7, learner control affects the modality effect in complex ways. Hatsidimitris and Kalyuga (2013) point out that unfamiliar learner control tools can even interfere with learning: the effort needed to operate these tools (e.g., moving the mouse, finding the correct playback point) can draw working memory away from processing the content, increasing extraneous cognitive load.

Unfamiliar devices. Another issue of ecological validity is the learner's use of their own learning tools versus those provided by the researchers. Learners may have years of experience with their own devices but none with ones used in research. It could be, for example, that the learner controls of videos require a higher level of technical skill with the device compared to the illustrated text, which could weaken the modality effect with unfamiliar researcher-given devices. Regrettably, studies commonly do not tell if the learners used their own tools or researcher-given tools, which limits the assessment of generalizability.

Short materials. Some laboratory experiments have used very short materials, which do not resemble classroom learning materials. Indeed, Rummer et al. (2011) propose that the modality effect for short materials may be due to differences in sensory memory, not working memory. Therefore, experiments conducted with short materials likely measure the micro-level modality effect (Section 2.1) with limited relevance for classroom learning.

Neglecting pictures. In reading research, visuals are frequently regarded as confounding variables that must be controlled (e.g., Hare, 2024). For example, in Clinton-Lisell's (2022) meta-analysis of reading and listening comprehension, materials with pictures were systematically removed from the review as the presence of pictures was seen as a confounding factor. However, this approach diminishes the studies' generalizability to primary education, where materials have abundant pictures. In general, there is a limited focus on pictures (e.g., Brysbaert, 2019), despite their central role in multimedia learning (Mayer, 2014).

2.10.3 Ecological validity of the tests

In contrast to settings and materials, the learning assessments in earlier modality effect studies have largely avoided ecological validity concerns. Question formats often align with typical educational practices, such as multiple-choice and short written answers. Similarly, cognitive load has often been evaluated using self-report scales, which resemble the self-assessment questions found in educational workbooks. However, there are also deviations from classroom practices.

Studying based on retention. Most learning experiments involve a study phase followed by an immediate posttest based solely on memory. However, in primary education, children are rarely expected to complete exam-style assessments immediately after being introduced to new topics. Administering such tests can increase stress and waste time by requiring students to recall content they have not yet had the opportunity to learn. Instead, in a common primary school classroom situation, children first complete exercises with the material available (a type of generative active processing, Fiorella & Mayer, 2021; Fyfield et al., 2022). Only after such tasks, and commonly after a delay, there is an exam where the children answer questions without the material.

Physical measures such as heart rate monitors, skin conductance sensors, and pupil diameter recordings, as well as eye-tracking systems, are rarely, if ever, part of classroom learning. Therefore, they can jeopardize ecological validity. For instance, being informed that their gaze is being tracked can make children feel observed. This may shift their attention away from the material or create pressure to perform. Wearable eye-tracking glasses may distract learners and draw unwanted attention from peers, especially when the devices appear unusual or humorous to children. Similarly, heart rate monitors remind the students that the lesson is not just a part of everyday school activity. In short, these measures alter the learning situation and may change the very outcomes they are intended to assess.

Dual tasks. Some studies have used dual tasks for cognitive load measurement, which can significantly alter the primary task of learning. This is exemplified by two studies using dual tasks (Brünken et al., 2002, 2004), in which learners had to monitor a flashing light or detect beeps while studying. These tasks confound the modality effect by interfering with reading and listening to a different degree. Reading and visual monitoring tasks both demand visual attention, while listening and auditory monitoring tasks both demand auditory attention.

Interrupting learning with repeated assessments. In some studies, cognitive load has been measured repeatedly during learning using self-rating scales, which can significantly disrupt the learning process. For instance, in the materials used by Tabbers et al. (2001, 2004), every other slide was dedicated to measuring cognitive load. This approach introduces pauses in the learning sequence, which may reduce cognitive load simply by offering breaks (Cheon et al., 2014; Xie & Salvendy, 2000). Rating tasks in the learning phase can also interrupt working memory processes, limiting learners' ability to integrate information. Furthermore, such procedures may introduce modality-specific biases, as only narrations require additional actions to pause or play between ratings.

3 Studies and Research Questions

Today, producers of educational materials commonly offer both illustrated text and video versions of their content, and most teachers regularly use educational videos (Hyvönen et al., 2018; Pattier, 2021). However, empirical research has not kept pace with this development. Most studies on the modality effect have been conducted with higher education students in laboratory settings, often with materials that do not offer learner controls. This context differs significantly from how children study in school today, and thus prior findings do not reliably generalize to primary education.

The three empirical studies of this dissertation address these gaps. All studies were conducted in Finnish primary schools and used crossover experiments to compare learning from videos and illustrated texts. The aim was to understand how these two learner-controlled multimedia formats support the learning of fifth- and sixth-grade students in classroom contexts. Rather than attempting to isolate a single mechanism causing the modality effect, the studies assess the relationship of multiple underlying factors, such as cognitive load, studying rate, and individual differences.

The theoretical background (Section 2) outlined multiple explanations for why narrated multimedia leads to reduced cognitive load. Cognitive load theory further posits that this reduced cognitive load explains improved learning results. However, this has not been empirically tested, especially in primary school classroom contexts. In particular, most studies that appeal to explanations from cognitive load theory have not statistically assessed whether cognitive load predicts learning outcomes (e.g., Huang et al., 2016; Inan et al., 2015; Liu et al., 2021), and some did not measure cognitive load (e.g., Knoop-van Campen et al., 2018, 2019; Segers et al., 2008).

Study I addressed this gap by exploring the modality effect in classroom settings with a crossover experiment. In the experiment, fifth graders learned science topics from both illustrated texts and videos. They completed pre-, post-, and delayed tests to assess cognitive load and learning outcomes. Individual difference factors were quantified by the combination of the within-subjects design with linear mixed-effects analysis. Furthermore, the direct effect of cognitive load on each learning outcome was assessed.

Past research on the modality effect, including Study I, has predominantly focused on learning outcomes without collecting information about the learning process, which limits the interpretation of the results. Study II aimed to assess alternative process-based explanations for the modality effect. It analyzed screen recordings that were captured on children's iPads during the learning phase of Study I. The analysis examined children's studying rates and strategies when they learn from videos and illustrated texts, and how these variables predict learning outcomes. In addition, the study assessed the influence of presentation characteristics on studying rate, to better understand how the design of the learning materials affects the learning process.

Prior research often explains that poor decoding increases cognitive load in written materials, leading to modality effect. However, commonly these studies have not measured reading skills or assessed their direct impact on cognitive load or the modality effect. Study III addressed this gap by replicating Study I, except for two key changes. First, it measured decoding ability and reading comprehension and analyzed how they interact with learning from videos versus illustrated texts. Second, it allowed students access to the instructional materials during the posttest, reflecting a type of studying more common in primary school classrooms but rarely implemented in modality effect research. This also enabled a comparison of studying based on retention versus retrieval between Studies I and III.

This dissertation focuses on three types of learning outcomes: retrieval, retention, and transfer. **Retrieval** is defined as the learner's ability to find and extract relevant information from the instructional materials in order to answer direct questions. The necessary information is directly presented within the provided materials. **Retention** refers to the learner's ability to respond to direct questions based on information stored in long-term memory. This was measured immediately after and one week after the initial study session, without access to the materials that originally contained the answers. **Transfer** involves the learner's ability to apply the knowledge gained to new situations by answering open-ended questions that require understanding of the studied content. The materials did not include direct answers to the questions.

In addition to these learning outcomes, **cognitive load** was measured using a direct subjective measure in the posttest: a single-item mental effort scale (Paas, 1992). Furthermore, **efficiency** metrics were calculated by combining learning performance with cognitive load. Note that learning time was controlled, and thus the three main factors contributing to efficiency in classroom learning were accounted for.

The studies also analyzed process data: screen recordings were used to assess **studying rate** (how quickly students progressed through the material in initial study) and **studying strategies** (in what order and pace did they progress in the material). These behavioral measures are particularly important given the shift toward learner-controlled materials in modern classrooms. Finally, the predictive effects of presentation characteristics on studying rate were assessed. **Text length** refers to the number of text characters on a slide. **Picture types** used in a slide were either diagrams or photographs or both. **Slide number** refers to the position of a slide in the presentation.

The topics of learning were climate change and balance in nature. In this dissertation, the word **topic** refers to the two topics of study (climate change and balance in nature), whereas **domain** refers to the wider domain of science learning, to which both topics belong.

Research questions are presented in Table 1. The broad dissertation-level questions directed the overall goals of the research, whereas the specific questions in each empirical study addressed these goals within the practical scope of each article.

Table 1. Research questions.

	Research questions
Dissertation	1. How does learning from videos versus illustrated texts affect learning outcomes, cognitive load, and the learning process when children study expository science content using modern learner-controlled materials in ecologically valid classroom environments? 2. Can differences in learning outcomes between modalities be explained by differences in cognitive load? 3. What alternative factors could explain learning differences between modalities, such as individual differences, reading skills, studying rates, review strategies, or characteristics of the learning materials?
Study I	1. Does learning from videos and learning from illustrated texts produce different learning outcomes measured by retention, delayed retention, and transfer? 2. Does learning from videos and learning from illustrated texts produce different cognitive loads or instructional efficiencies? 3. What is the relationship between cognitive load and learning outcomes? 4. What is the role of individual differences in the modality effect?
Study II	1. Does studying rate differ when learning from videos compared to illustrated texts? 1.1. Does (a) text length, (b) picture type, or (c) slide number affect the studying rate? 1.2. Does the studying rate predict learning or cognitive load? 2. What strategies do students use to review illustrated texts and videos? 2.1. Do the review strategies predict learning or cognitive load?
Study III	1. To what extent does the decoding ability and the reading comprehension affect retrieval, retention, transfer, and cognitive load when studying videos versus illustrated texts? 2. Does studying based on retrieval from videos versus illustrated texts produce different outcomes in retrieval, retention, transfer, and cognitive load?

4 Methods

The three studies of this dissertation investigated the modality effect with crossover experiments, in which fifth- and sixth-grade students studied videos and illustrated texts. The materials were presented in ways that closely mirrored classroom learning, while ensuring that the two modality conditions remained comparable. Because all participants learned from both modalities and mixed effects models were used in the analyses, the research could quantify individual differences. Moreover, since the different modalities were learned from simultaneously in the same environment, the study design also controlled for environmental factors.

Table 2 presents an overview of the methods used in this dissertation, and the following sections provide a more detailed discussion of key methodological considerations.

Table 2. Overview of the methods of this dissertation.

Study	Sample	Measures and materials	Procedures
I	- 54 fifth graders across three classes.	- Pre-, post-, and delayed tests assessing prior knowledge, retention, and transfer through multiple-choice, pictorial, and open-ended questions. - 9-point self-rated cognitive load scale. - Revised multimedia materials from Knoop-van Campen et al. (2018).	- Within-subject design with two science topics (climate change, balance in nature). - Random assignment to video or illustrated text conditions. - Learning sessions and assessments conducted in classrooms using personal tablets.
II	- 97 screen recordings from Study I participants.	- Screen recordings captured studying rate, studying strategies, navigation actions. - Learning outcome and cognitive load data from Study I.	- Screen recordings obtained via participants' tablets during learning sessions. - Frame-by-frame coding of slide durations and navigation actions using BORIS software. - Categorization of review strategies based on studying rate and order.
III	- 109 fifth- and sixth grade students across six classes	- Measures from Study I, except that posttest retention questions were retrieval questions because the material was available. - Reading comprehension and decoding ability tests. - Materials from Study I.	- Replication of Study I with the following additions: (a) assessments of reading skills and retrieval, (b) posttest conducted with material available.

4.1 Ethical considerations

The research complied with the ethical principles of research with human participants (Finnish National Board on Research Integrity TENK, 2019). Study III, which used largely the same research procedure as Study I and II, was approved by the University of Turku's ethics board.

Research permits were obtained from municipal representatives, school principals, and class teachers. Written informed consent was provided by participating children and their legal guardians. Participants were informed of the study design. Participation was voluntary and participants could withdraw from the study at any time without consequences. Participants were not compensated for participating. Participants and teachers were notified when the findings were published.

The study did not involve physical, psychological, social, or financial risks. All activities occurred during regular school hours and aligned with the national curriculum. The researcher, a qualified teacher, instructed the lessons together with class teachers. Class teachers reviewed learning tasks and questions for appropriateness. Teachers and the researcher assisted with instructions when necessary. Those who did not wish to participate attended the same lessons, but their data was not collected for research. This approach ensured that the situation resembled a lesson instructed in part by a visiting guest teacher.

Data collection included pretests and posttests, surveys, and screen recordings. Each student received an ID code for use in the study. The researchers did not know at any point which student had which ID code. Because the lessons were part of normal schoolwork, students who did not participate in the research returned their tests to their teacher under their own name. Students who participated returned their tests to the researchers using only the ID code. If requested by the teacher, participating students shared their ID code with the teacher, who could therefore identify which student corresponded to each ID. The researcher later returned the scanned tests to the teacher if the teacher wished to keep copies. The surveys and screen recordings were returned to the researchers and were not shared with the teachers. The ID code included in each data source allowed the researchers to link data from the same participant without knowing the participant's identity.

The data are considered pseudonymized because the teacher could know the identity corresponding to each ID, even though researchers did not have access to this information. The collection of personal data was minimized: the pupil's and guardian's names and signatures were required only on consent forms. These records were not linked to the research data or ID codes at any point. No personal or sensitive data were collected. Screen recordings captured only interactions with the learning materials, without audio.

All data were handled confidentially and stored on a secure, password-protected server accessible only to the research team. Paper tests were scanned and then destroyed. Consent forms were stored in paper format and not digitized. The legal basis for processing personal data was the public interest in scientific research, as outlined in the GDPR. A separate data protection impact assessment was not required because identifying data were used only for consent purposes.

4.2 Participants

In this dissertation, all participants were students attending primary schools in Finland where Finnish was the language of instruction. Schools were selected using convenience sampling from public primary schools in a local suburban area in Southwest Finland. Private schools, special education schools, and teacher training schools were excluded. The representativeness of the sample is increased by the fact that approximately 99% of lower primary school students in Finland attend public mainstream schools (Official Statistics of Finland, 2021), instead of special or private schools. Furthermore, after accounting for individual-level variance, school- and class-level variance was minimal, supporting the generalizability of the findings to other similar schools and classes.

Within the participating schools, special education classes, and classes where the primary language of instruction was not Finnish, were excluded. Within the participating classes, students who needed external support (e.g., translation) to understand instructions in Finnish were also excluded.

Study I collected data from three fifth grade classes in one school. The 54 participating children (26 boys, 28 girls) were on average 11.9 years old (SD = 0.4).

Study II used screen recordings from the same group of students in Study I. The recordings captured students' interaction with the digital learning materials during lessons.

In Study III, data were collected from three schools. In each school, one fifth-grade teacher and one sixth-grade teacher were invited to participate. If a teacher declined, another teacher from the same grade level was invited to ensure balance across grades. The final sample consisted of one fifth-grade and one sixth-grade class in each of the three participating schools. The participating 109 children (54 boys, 48 girls, 7 undisclosed) were on average 12.3 years old (SD = 0.6).

4.3 Data collection and procedure

All three studies used the same confounded factorial design (Kirk, 2009). In every class, two 45-minute science lessons compared learning from an illustrated text with learning from a video. Children were randomly assigned to group A or group B. In lesson 1, group A studied climate change with an illustrated text and group B with a video; in lesson 2, group A studied balance in nature with a video and group B with an illustrated text. Each lesson began with a pretest, followed by studying the material, and ended with the posttest. Delayed retention was assessed one week later.

In Study I, the studying phase was 12 minutes, during which the children had freedom to study the material at the pace and in the order of their choice. This was roughly twice the duration of each narration, giving learners in both conditions enough time to study and review the presentation. After that, they had approximately 25 minutes to complete the posttests without access to the learning material. For Study II, screen recordings were captured from all participants to allow for further analysis of how they engaged with the content.

Study III replicated the design of Study I, with two key differences: (a) a separate 45-minute lesson was conducted within a month of the experiment to assess decoding ability and reading comprehension, and (b) the children completed the posttests with access to the learning material. The posttests were distributed after the children had studied the material once from start to finish. The combined time for studying and posttest was about 35 minutes in each of the two lessons. Figure 1 illustrates the sampling and data collection procedure in Study III.

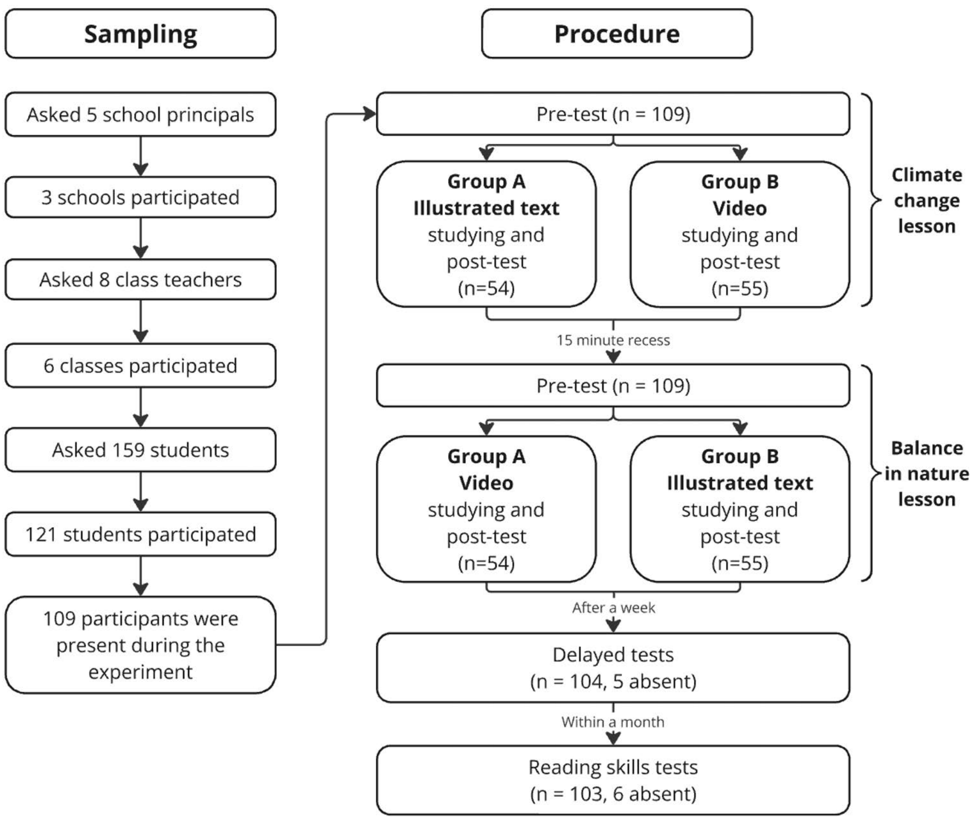

Figure 1. Sampling and data collection procedure in Study III. Figure from Haavisto et al. (2025).

Study II analyzed screen recordings that were collected from tablet computers during the 12 minutes learning phase in Study I. These recordings showed children's pace and order of progression in the illustrated texts and videos, as well as what navigation actions were used. A total of 97 screen recordings were collected. In the first lesson, 21 recordings were captured for the video condition and 26 for the illustrated text condition. In the second lesson, 26 recordings were collected for the video condition and 24 for the illustrated text condition. Two Study I participants did not participate in Study II, one participant was absent during the second lesson, and six recordings were lost due to technical problems, such as the recording failing to start.

The methods used in this dissertation were chosen to maintain the natural flow of the lessons. The experiments were carried out during regular school hours. All learning and testing took place in the students' own classroom, with students sitting at their own seats, with their classmates present. The children studied the learning materials on their personal tablet computers, which they had used regularly for at least an entire academic year in each studied class. Teachers were present during the lessons to maintain classroom order and to help with the intervention procedure. However, they were instructed not to offer any information related to the studied topics.

The testing procedure was also chosen to ensure ecological validity. Tests were completed on paper; a format commonly used in classroom assessments in conjunction with studying on tablets. Cognitive load was self-assessed in the posttest, reflecting self-assessments commonly found in textbooks.

The timing of the lessons was designed to mirror classroom routines. As customary in Finnish schools, each lesson lasted 45 minutes and was followed by a 15-minute recess during which the children played in the school yard. The time allocated for studying each material limited, reflecting lesson-length. The delayed test was after one week, which tested for knowledge retention. This measure is relevant since it matches what students would retain when their schedule has two hours of science lessons once per week.

4.4 Learning materials

The same learning materials were used in the three studies. They consisted of paired versions of the same content: an illustrated text and a video for each of the two science topics, climate change and balance in nature (Figure 2). Both formats were adapted from materials developed by Knoop-van Campen et al. (2018, 2019). These materials were chosen out of those used in multimedia learning research because they resembled multimedia materials commonly used in Finnish primary schools the most with respect to style of expository text, pictures, and length. The texts were translated into Finnish and expanded by about one-third to fit the available lesson time. Revisions were completed in consultation with a participating teacher to ensure that the materials reflected those typically studied in primary education. The learner controls were considered familiar, the language and difficulty level were appropriate, and the content was novel for the participants.

For the two topics, the materials contained 1 383 words and 39 pictures in total, comprising 14 diagrams and 25 photographs. The climate change materials included 474 words and 18 pictures arranged on 16 slides and the video lasted 5 min 55 s. The balance in nature materials contained 466 words and 21 pictures on 15 slides and the video lasted 5 min 37 s. Spoken narration in the videos was presented at 78 words per minute in Finnish (for reference, English translations were 121 words per minute). In both topics, the illustrated text was studied as a Google Slides presentation in which text and pictures appeared side by side, and the video was studied on YouTube with the same text narrated instead of written. In the video condition, illustrations were scaled to fill the screen to avoid blank spaces where text had been removed.

Navigation options in both conditions reflected those commonly used in modern multimedia learning and were thus familiar to learners. The videos included hybrid learner control with continuous playback and standard navigational options (e.g., pause, rewind, fast-forward, and a timeline slider), while the illustrated text allowed free navigation forward, backward, or to any slide.

Every presentation ended with an additional review slide that listed the topics covered in the material and provided additional instructions. In Study I, the review slide informed learners that they still have time to review any sections of the material. In Study III, the review slide informed learners to raise their hand to receive the posttest to be completed with the material.

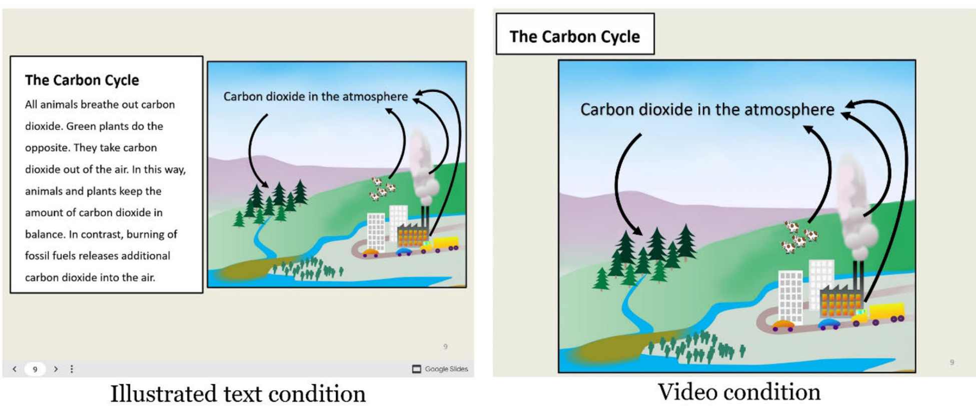

Figure 2. An example of illustrated text and video materials. The same text that was written in the illustrated text was spoken in the video. Figure from Haavisto et al. (2023).

4.5 Measures

All studies used the same revised and translated domain knowledge tests based on Knoop-van Campen et al. (2018) to assess retention, delayed retention, retrieval, and transfer. In collaboration with a teacher from a participating class, the test questions were revised to match the translated learning materials and to reflect primary school testing conventions that were familiar to the participants. The same domain knowledge questions were used in the pretest, posttest, and delayed test, except that the pretest included only a subset of the questions.

Retention and **retrieval** were measured with multiple-choice, pictorial, and closed-ended written-answer questions assessing factual information from the learning materials. For example, retention questions included prompts such as "Write four examples of fossil fuels." and "Draw four arrows in the image. Show with the arrows, which direction carbon dioxide transfers between: 1. Air and forest. 2. Air and factory. 3. Air and cars. 4. Air and cows.". The test items were designed so that some questions could be answered only from the pictures, some only from the text, and some from both, ensuring that both words and pictures were instructionally relevant. In Study III, the posttest questions measured retrieval, as the material was available during posttest, unlike in Study I, where the same questions measured retention.

Transfer was measured with open-ended questions that required the learner to apply the studied information to novel contexts. A sample transfer question was: "Imagine that researchers have invented three machines, which help prevent climate change. Write below what these machines do and how it helps prevent climate change."

Each question was divided into one to four scored items (e.g., "Name two greenhouse gases" contained two items), with responses scored as 1 point for correct and 0 points for incorrect answers. Partial scoring was used in a subset of items. Total scores reflected the number of correct items.

In all studies, participants self-assessed their level of **cognitive load** during studying using Paas's (1992) nine-point mental-effort scale. The self-assessment item was placed as the first question of the posttest. The scale ranged from very very low effort to very very high effort. The scale was explained thoroughly to participants, as in the original implementation of this scale (Paas, 1992). In Study I, **efficiency metrics** were calculated by subtracting the standardized cognitive load score from each standardized learning outcome score (retention, delayed retention, or transfer) and dividing the result by the square root of two (Paas & Van Merriënboer, 1993). As the cognitive load scale measured participants' mental effort during learning, the efficiency metrics represent adapted instructional efficiency (Van Gog & Paas, 2008).

In Study II, these same knowledge and cognitive load data collected in Study I were used to assess the relationships between performance measures and studying process variables.

In Study III, **reading comprehension** was measured using the Finnish Standardized Reading Test (Lindeman, 1998), which included two informational texts with 12 multiple-choice questions each. Children were allowed to refer to the texts throughout the test. **Decoding ability** was assessed with a validated test battery (Nevala & Lyytinen, 2000), which consisted of four timed tasks: (a) identifying individual words in a continuous text without spaces, (b) separating syllables within words, (c) distinguishing real words from pseudowords, and (d) detecting spelling errors.

4.6 Analyses

All the analyses that required within-subjects variables (i.e., the repeated measures of student performance across varying modalities and topics) were conducted using mixed-effects models. Separate models were constructed for each dependent variable. All mixed-effects models included random intercepts for participant ID to account for individual differences and within-subject dependencies. Fixed effects included modality (video or illustrated text) in Studies I and III, and also decoding and reading comprehension and their interaction with modality in Study III.

Statistical significance is reported using a threshold of $p < .05$. Effect sizes are reported in terms of conditional R^2 (R^2_{con}, variance explained by fixed and random effects) and marginal R^2 (R^2_{mar}, variance explained by fixed effects; Nakagawa & Schielzeth, 2013). They are interpreted using guidelines for partial eta squared: R^2_{con} or $R^2_{mar} > .01$ indicates a small effect, R^2_{con} or $R^2_{mar} > .04$ a medium effect, and R^2_{con} or $R^2_{mar} > .16$ a large effect. While R^2_{mar} is the combined variance explained by all fixed effects, ΔR^2_{mar} denotes the variance explained only by the effect in question. The difference between conditional and marginal R^2 ($R^2_{con} - R^2_{mar}$) represents the variance explained by random effects, which in this dissertation corresponds to variance explained by individual differences.

Mixed-effects models allowed the efficient use of all available data across repeated measurements, including cases with missing values due to student absences. Mixed-effects models are also robust to imbalances in comparison group sizes.

All analyses were conducted using R software (R Core Team, 2024) with the R packages lme4 (Bates et al., 2015) for model construction, performance (Lüdecke et al., 2021) for calculating model fit indices and effect sizes, clubSandwich (Pustejovsky & Tipton, 2018; Pustejovsky, 2022) and sandwich (Zeileis, 2004) for heteroskedasticity-robust statistics, and simr (Green & MacLeod, 2016) for power analysis. Screen recordings were examined using BORIS software (Friard & Gamba, 2016).

An overview of the statistical tests used in each study is presented In Table 3. The subsequent sections detail how screen recordings were preprocessed and analyzed in Study II (Section 4.7), and how validity and reliability were assessed (Section 4.8).

Table 3. Overview of the statistical tests of this dissertation.

Statistical test	**Study and purpose**
Linear mixed-effects models	Study I: To analyze the effect of modality and cognitive load on retention, delayed retention, transfer, and the effect of modality on cognitive load and efficiency. Study II: To analyze the predictive effect of studying behaviors on Study I performance measures. Study III: To analyze the effect of modality on retrieval, delayed retention, cognitive load, as well as interaction effects between modality and reading skills.
Chi-squared tests, likelihood ratio tests	Studies I, II, III: To compare model fit with and without specific fixed effects.
Paired t-tests	Study I: To compare pretest scores with posttest and delayed test scores.
Cronbach's α	Studies I, II, III: To assess internal consistency reliability of test scales.
Intraclass correlations	Studies I, II, III: To assess variance at participant and classroom levels, guiding inclusion of random effects.
Inspection of residual and quantile-quantile plots	Study I: To confirm that residuals are approximately normally distributed and homoscedastic.
Heteroskedasticity-robust standard errors	Studies II, III: To provide robust estimates without assuming homoscedastic or normally distributed residuals (White, 1980).
Stepwise regression	Study II: To analyze effects of slide-level predictors on study time.
Pearson correlation	Studies I, II, III: To examine relationships between performance measures.
Conditional and marginal R^2	Studies I, II, III: To report effect sizes in mixed-effects models.
Variance inflation factors	Studies II, III: To confirm acceptable levels of multicollinearity.
Case deletion diagnostics	Studies I, II, III: To assess the degree to which the results were sensitive to potentially influential observations (Fox, 2020).
Welch's t-test	Study III: To confirm that practice effects were similar between modalities.
Linear regression	Study III: To confirm no pretest differences between the randomized experimental groups.
Simulation-based power analyses	Study III: To assess the minimum effect size detected with tolerable levels of type two errors.

4.7 Screen recording analysis and measures

In Study II, the screen recordings were analyzed frame by frame. Coding began when the learner progressed from the title slide to the first multimedia slide and ended when they exited the learning material. For every frame where a slide was visible, its sequential number within the presentation was recorded. This made it possible to determine both the order in which the slides were accessed and the amount of time spent on each one.

The coding process was split into two phases. The **initial pass** began when the participant first entered the first multimedia slide and ended at the first entry into the final slide of the presentation, referred to as the review slide. The **review** phase covered the period from the first entry into the review slide until the participant exited the material. Following the approach used in Mason et al. (2013), data from the initial pass was used to assess the studying rate for new content, while data from the review phase was analyzed to identify review strategies.

All studying rate analyses were based on the durations calculated from the frame-by-frame coding. To control for topic-specific differences in narration length, each studying duration was expressed as a proportion of the corresponding video's total duration. The final studying rate was calculated by dividing the number of words and pictures by the studying duration in minutes. This resulted in metrics of words per minute (wpm) and pictures per minute (ppm).

The materials were in Finnish, with an average word length of 7.55 characters. To make the results easier to interpret and compare with previous research, word counts and wpm values were based on the English translations of the texts, which had a mean word length of 4.84 characters. The original Finnish values can be retrieved by dividing the reported figures by 1.561. This conversion method is supported by Brysbaert (2019), who showed that differences in reading speed across languages and texts can be explained by word length. Liversedge et al. (2016) also found that translating between Finnish and English does not affect sentence-level reading times.

To examine how **slide characteristics** affected study time during the initial pass, the data were aggregated at the slide level. Three slide-level predictors were considered. First, the character count of the text on each slide was used. Second, slide number was analyzed to detect whether students spent more time on slides appearing at specific points in the presentation. Third, a variable was created for picture type: slides with diagrams were coded as 1, those with photographs as 0, and two slides with icons or a mix of diagram and photograph as 0.5. Stepwise regression was used to test the effects of these predictors. Each model containing one of the predictors was compared to a baseline model using F-tests, and predictors were added in order of their F-statistics. Interaction effects were also examined. Topic accounted for less than 0.5% of the variance in study time, indicating strong consistency between the two science topics.

Three **categories of studying rate** were defined based on thresholds of 0.4 and 2.5 times the narration rate of the corresponding video. These thresholds corresponded to less than 47 wpm + 1.3 ppm or more than 295 wpm + 8.3 ppm and were informed by Kuperman et al.'s (2021) definitions of reading that is either too fast to support comprehension or so slow that it may indicate technological problems or distraction.

Studying strategies were determined by categorizing studying rate and order. Scanning was coded for all slides where the time spent was less than 0.4 times the video's duration for that slide. Slowing was coded for slides where the time spent was more than 2.5 times the duration. For the remaining data, studying behavior was further classified as forward linear when part of a forward sequence involving at least two slides, backward linear when part of a backward sequence, or isolated when not part of any linear sequence.

4.8 Validity and reliability

4.8.1 Reliability analyses

In all studies and for both lesson topics, posttest and delayed test retention, retrieval, and transfer **internal consistency reliabilities** measured by Cronbach's alphas were in the range .67–.83, which were deemed sufficient (Taber, 2018). After ensuring reliability, scales of both topics were combined, further increasing reliability.

Participant-level **intraclass correlations** (ICCs) were high for all dependent variables (ICCs = .424–.546), justifying the inclusion of participant random effects in all mixed-effects models to account for within-subject dependencies. In contrast, class-level ICCs were low across all response variables (ICCs < .021 in Study I; .013–.083 in Study III), indicating minimal clustering at the classroom level. Therefore, classroom-level random effects were not included in the final models.

Multicollinearity was assessed using variance inflation factors for every model with multiple fixed effects. Models without interaction terms showed variance inflation factors below 1.4, indicating minimal multicollinearity, and models with interaction terms remained within acceptable limits (variance inflation factors below 2.55; O'Brien, 2007).

In Study I, a randomly selected 20% subset of the pretest, posttest, and delayed tests for both topics was independently scored by a second rater. **Inter-rater reliability** was evaluated using two-way mixed, absolute agreement, single-measures intraclass correlations, all of which were excellent (> .90, Koo & Li, 2016). This indicates that the coding process introduced minimal measurement error. Any differences between the two sets of scores were reviewed and all responses were rechecked for inconsistencies. Study II used the same learning and cognitive load data as Study I. In Study III, the scoring criteria and the scorer were the same as in Study I, and therefore, it was not re-evaluated in Study III.

In Study III, a power analysis of 1000 simulations was conducted for each model, showing sufficient **statistical power**, detecting Cohen's $d \geq 0.3$ effect sizes with 87.2% probability on average.

4.8.2 Prior knowledge and other individual differences

In Studies I and III, the pretest questions for both lesson topics showed low reliability (alphas < .60), and thus prior knowledge was not included in the models. The same questions yielded acceptable reliability when used in the posttest, and therefore the low pretest alpha values were not likely due to poor item quality. Instead, they were likely due to a floor effect as the pretest scores were very low. While these low pretest scores did not reliably capture differences in prior knowledge, they do show that prior knowledge was low overall. This was expected, as the lesson content was specifically designed to present new information to the participants.

Instead of relying on pretests to control for prior knowledge, mixed-effects models with random effects for participants were used to account for all individual differences, including prior knowledge. This approach is more robust than controlling for prior knowledge alone. Furthermore, random assignment of participants to conditions balances individual differences between comparison groups.

The treatment of prior knowledge in the analysis followed the same logic as for age and gender variables: these variables were measured and reported to inform generalizability but were not included in the models. This is because including both random effects and specific individual difference variables would partial out variance from the random effect, reducing statistical power.

4.8.3 Practice and testing effects

In crossover studies, a washout period typically follows the first treatment (i.e., the first lesson in this dissertation) allowing for the effect of the first treatment to diminish, making the starting point of the second treatment (i.e., the second lesson) more comparable. However, a washout period is not feasible in studies of learning due to the extensive time required to fully forget previously learned content. Therefore, this dissertation used two distinct topics in the interventions to reduce order effects such as practice effects.

Using different topics required different tests and materials, which led to differences in score means and variances. To address this, all measures of retrieval, retention, delayed retention, transfer, and cognitive load were standardized separately for each lesson before being combined for within-subjects analyses. This approach accounted for topic-specific differences, enhanced statistical power, and supported the goal of assessing how learning would generalize across science topics rather than compare the specific studied topics.

Within each science lesson, identical questions were used in the pre-, post-, and delayed tests, which introduces a testing effect. However, this was not a concern because the randomly assigned modality conditions (video and illustrated text) were conducted simultaneously in the same classroom and followed the same procedures. Therefore, any practice or testing effects were balanced between conditions. Furthermore, participants were not informed that the same questions would be used again, and they did not receive feedback from testing before all testing was completed. Note that transfer was assessed only in the posttest. Lastly, answering questions reflects a common learning method in primary education.

4.8.4 Cognitive load measurement

The chosen cognitive load measure, the single-item mental effort self-rating scale (Paas, 1992), is widely used in cognitive load research (Mutlu-Bayraktar et al., 2019) and is considered a reliable and valid indicator of overall cognitive load (Chen et al., 2016; Gopher & Braune, 1984; Paas et al., 1994, 2003; Sweller et al., 2019; van Gog & Paas, 2008). However, as cognitive load measurement is a topic of active debate (e.g., Anmarkrud et al., 2019), a further justification of the choice of cognitive load in this dissertation follows (see also Sections 2.3 and 2.10.3).

Schmeck et al. (2015) found that self-assessed cognitive load after a task (i.e., delayed rating) is consistently higher than the average of ratings collected during the task. Van Gog et al. (2012) reported similar results. However, delayed rating did not inflate other variables like interest or motivation (Schmeck et al., 2015). This pattern supports the hypothesis that pauses introduced by cognitive load questions during a task may reduce the very load they aim to measure. During a demanding task, stopping to answer a question offers a pause, which could lower the cognitive load. Delayed ratings avoid this issue. Lastly, in this dissertation, potential upward bias in delayed rating would affect both conditions equally, and thus the difference in cognitive load between modalities still reflects a real effect.

A criticism of self-rating scales is their subjective nature: participants may interpret the scale differently or respond differently to similar experiences (Jiang & Kalyuga, 2020). These individual differences were accounted for by the within-subjects design and the inclusion of random effects for participants in the statistical analysis. In simpler terms, each participant’s cognitive load was compared to their own cognitive load in the other condition.

Lastly, some prior studies on the modality effect (e.g., Leahy & Sweller, 2011; Wong et al., 2012) have criticized the use of self-ratings with children due to negative item bias and extreme response bias (see Chambers & Johnston, 2002; Marsh, 1986). These concerns did not apply in this dissertation: the scale does not include negatively worded items, and extreme responses were rare. In addition, Ayres (2006, Experiment 2) found subjective ratings to be valid in children of similar age.

5 Main Findings and Theoretical Implications

An overview of the studies is presented in the abstract. This section discusses the main findings of the studies, along with descriptive statistics, statistical significance levels, and effect sizes (Section 4.6). All descriptive statistics reported for groups are means unless otherwise stated. Full results and more detailed comparison with earlier empirical studies are presented in the attached articles (Haavisto & Jaakkola, 2025; Haavisto et al., 2023, 2025).

The three studies in this dissertation investigated the modality effect in primary school classrooms, using learner-controlled materials. The goal was to identify the mechanisms that underlie the modality effect in learning outcomes and cognitive load.

In Study I, 54 fifth-grade students participated in a crossover experiment where students learned two science topics (climate change and balance in nature) using both illustrated texts and videos. The study examined the effect of modality on immediate retention, delayed retention, transfer, and cognitive load. In addition, the relationship between cognitive load and learning outcomes was analyzed.

Study III replicated Study I with a larger sample of 109 fifth- and sixth-grade students, with two additional modifications. First, posttest retention questions were answered with the learning materials available, turning them into retrieval questions. This adjustment aimed to closely resemble the classroom context and to compare studying based on retention versus retrieval from materials. Second, students' reading comprehension and decoding ability were measured to determine whether reading skills account for the potential benefits of videos.

Study II is introduced in Section 5.8.

This dissertation focused on three major explanations based on cognitive load theory (Section 2.4) that can favor either narrated or written formats: (a) decoding written text increases unnecessary load, which reduces learning from written multimedia; (b) split attention increases extraneous load, which reduces learning from written multimedia; and (c) transient information must be held in working memory, increasing extraneous load, which reduces learning from narrated multimedia.

There are three key perspectives that guide the interpretation of the results. First, a positive modality effect suggests that the effects that favor videos (e.g., decoding efficiency, reduced split attention) outweigh negative factors. In contrast, a reverse modality effect (favoring text) would suggest that the advantages of illustrated texts, such as less transient information, outweigh the benefits of videos. In addition to these primary explanations based on cognitive load theory, alternative explanations (e.g., written text offers more efficient lookup of key facts, videos are studied faster) are explored that could explain the observed effects.

Second, the modality effect has often been interpreted through the lens of a single theoretical explanation, such as split attention in the case of a positive modality effect or transient information effect in the case of a reverse modality effect. However, there are multiple mechanisms through which either narrated or written multimedia could have an advantage, and none of these mechanisms have been ruled out. Therefore, they should all be taken into account, and the modality effect should be considered as the combined effect of multiple simultaneous positive and negative effects. Even in cases where no modality effect is found, it does not imply that the underlying processes are absent. For example, the transient information effect may hinder learning from videos, while the decoding demands of reading may simultaneously hinder learning from texts. If these opposing effects are roughly equal in strength, the result is no observable modality effect, despite meaningful differences between cognitive processes.

Third, some mechanisms are likely to be stronger in effect for specific situations. For instance, prior studies using short stimuli in system-paced conditions under high performance pressure may frequently induce cognitive overload, thus increasing the explanatory power of cognitive load theory (Section 2.4). In contrast, the underlying effects in the present dissertation are likely different, as it studied classroom learning with longer learner-controlled materials.

5.1 Learning from illustrated texts and videos

Learning and forgetting occurred across the test phases, as expected, in both illustrated text and video conditions and across studies. In Study I, the children answered only 19.4% of the pretest questions correctly, indicating low prior knowledge. After studying, they got 59.8% correct in the same questions in the posttest. A week later, the result decreased to 50.6% in the delayed test. Post hoc tests confirmed significant learning in both conditions when comparing posttest and delayed test scores to the pretest scores (p-values $< .001$).

In Study III, scores on the same subset of questions reveal similar performances across test phases (Figure 3). Again, prior knowledge was low, with only 14.2% correct on the pretest questions. In the posttest, children achieved 77.0% correct answers when the material was available. A week later, the score regressed to 54.3% correct in a delayed test conducted without access to the material. Again, post hoc analyses showed large performance differences between test phases ($\Delta R^2_{mar} = .363$—.886, p-values < .001).

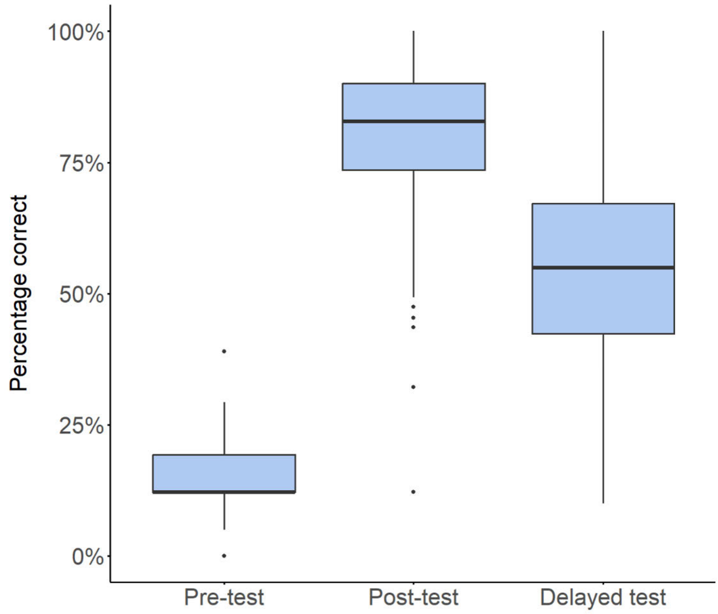

Figure 3. Performance differences in Study III pretest, posttest, and delayed test. Figure from Haavisto et al. (2025).

These results, and descriptive statistics in Table 4, show that learning occurred, which is important in the sense that both videos and illustrated texts are effective for classroom learning, whether they are studied with retention or retrieval testing. While these findings are hardly novel, what makes them more interesting is the comparison between studying based on retention testing (Study I) versus retrieval from materials (Study III). The questions in the posttests and delayed tests were the same in these studies, and prior knowledge differed only by 2.3% in matched items, supporting the comparison. However, since material availability was not manipulated in a single experiment, only descriptive findings are discussed.

Table 4. Learning and cognitive load in videos and illustrated texts.

Dependent variable	**Study I** (*n* = 51–54)				**Study III** (*n* = 104–109)			
	Video		**Illustrated text**		**Video**		**Illustrated text**	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Cognitive load	4.30	1.30	5.04	1.53	3.74	1.54	4.36	1.39
Posttest, retrieval	-	-	-	-	.779	.152	.760	.187
Posttest, retention	.645	.176	.567	.240	-	-	-	-
Posttest, transfer	.415	.205	.360	.223	.279	.242	.285	.252
Delayed test, retention	.574	.178	.482	.190	.598	.164	.549	.210

Note. Means are proportions of correct answers, except for cognitive load. Table is a combination of tables in studies by Haavisto et al. (2023, 2025).

In the posttest, Study III reported scores that were 26.3% higher with the material available, compared to Study I. This was expected, as answering the same questions by retrieving the answers from the material is easier than recalling the answers from memory. More notably, students in Study III also performed 8.3% better in the delayed retention test, where the material was not available in either study. This suggests that retrieving information from materials during the posttest helped reinforce learning. In contrast, if a student could not recall an answer in the retention test, they missed the opportunity to reinforce that knowledge. On the other hand, when students did remember the answer, both material availability conditions allowed reinforcement through recall. Once again, there might be simultaneous, opposing underlying effects. It is plausible that the absence of materials encourages students to invest more effort into recall, potentially resulting in stronger learning gains for successfully recalled elements, but this effect was weaker than the effect of being able to rehearse by retrieval from materials. These factors likely both depend on the proportion of successfully recalled answers as well as the effectiveness of the available navigation options.

While studying based on retrieval from materials was advantageous for delayed retention, it led to 26.4% lower transfer performance compared to studying based on retention. This may be due to two factors. First, students likely spent more time on the retrieval questions in Study III, leaving less time to complete the transfer questions that came later in the posttest. In contrast, the students likely proceeded to the next question more quickly if they did not remember the answer in the retention tasks of Study I. Second, when studying with the material available, the children may have tried to search for answers in the material, which was unhelpful since the transfer items were designed in such a way that the answers could not be found in the materials.

These results illustrate both the benefits and limitations of studying based on retention versus retrieval and exemplify that the learning process can be substantially different based on interacting factors such as material availability, test difficulty, test instructions, learning time, and question count. They also emphasize the importance of ecological validity (Section 2.10). A single change, such as material availability, can necessitate adjustments to other elements to facilitate learning. For example, when materials are not available, there is (a) no need to instruct students that transfer task answers are not in the material, and (b) less need to instruct students on allocating their time evenly across test questions.

Teachers typically provide such adaptive instructions, resulting in more effective learning and fairer comparisons between study formats. In contrast, controlled research often aims to minimize variation between conditions, potentially compromising the relevance of such comparisons. The present dissertation made this mistake by not adapting the instructions accordingly: retention-based studying was taken as a starting point and studying based on retrieval from materials was attempted to be implemented with just a single change, material availability. Such mistakes are common in multimedia research (Section 2.10.2).

5.2 Modality effect in cognitive load

In both Study I and Study III, the video condition led to significantly lower cognitive load than the illustrated text condition. In Study I, the modality effect was medium in effect size ($\Delta R^2_{mar} = .070$, $p < .001$). Study III replicated this finding with a similar effect size ($\Delta R^2_{mar} = .046$, $p < .001$). These results indicate a robust modality effect on cognitive load, favoring videos in authentic, learner-controlled classroom conditions. They are also consistent with cognitive load theory and prior research on the modality effect (Sections 2.1, 2.3, and 2.4).

Multiple theoretical explanations have been proposed for why videos might impose less cognitive load than illustrated texts. First, listening requires less mental effort than reading because it bypasses decoding written words. Second, videos avoid the cognitive effort caused by splitting attention between text and images, which can be viewed and listened to simultaneously. Third, learners' working memory functions more efficiently when information is presented through both auditory and visual channels. There are also explanations that suggest the opposite result. The transient information effect suggests that the temporary nature of videos imposes higher cognitive load, as learners must process and integrate information before it is quickly replaced in the automatically progressing video stream.

Note that despite the apparent contradiction, the observed modality effect is consistent with the transient information effect, as the increased load from transience could have been outweighed by simultaneous reductions in load from other factors.

With multiple potential underlying effects, none of the explanations can be ruled out based on the results. However, the results do suggest that the positive factors outweighed the negative ones. The decoding explanation is revisited in Section 5.7, where reading skills measurements facilitate stronger conclusions.

While differences in cognitive load between conditions were consistently lower in videos, it is also important to consider the absolute levels of load reported across conditions. In Study I, the children rated their cognitive load as 4.7 on a 9-point scale, corresponding to *neither low nor high effort*. In Study III, the average cognitive load was even lower, close to *somewhat low effort*. These moderate levels suggest that learners in both conditions likely had remaining cognitive capacity during the learning tasks. This is a crucial point because cognitive load theory is primarily concerned with avoiding overload. When overload is absent, reducing extraneous cognitive load, while still potentially beneficial, may have only a limited impact on learning outcomes. This discussion is continued in Section 5.5, after first discussing the relevant learning and efficiency results.

5.3 Modality effect in learning

In Study I, modality had a significant effect on retention scores. The videos outperformed the illustrated texts in posttest retention with a small effect size ($p = .001$, $\Delta R^2_{mar} = .030$) and delayed retention with a medium effect size ($p < .001$, $\Delta R^2_{mar} = .060$). In contrast, posttest transfer tasks showed no significant differences ($p = .093$).

The primary difference in Study III was that children had access to the learning material during the posttest, unlike in Study I. Despite this change, a similar pattern was replicated in the modality effects: videos outperformed illustrated texts in delayed retention ($\Delta R^2_{mar} = .016$, $p < .01$), while no effect was observed for transfer ($\Delta R^2_{mar} < .001$). Note that the modality effects were weaker with the material available, which suggests that allowing access to the material during testing might reduce the advantage of videos over illustrated texts. This was expected: videos might have a lesser advantage in studying based on retrieval from materials because illustrated texts facilitate easier location of relevant information, allowing the students to focus more on the content rather than the retrieval process (Section 2.4).

However, the retrieval performance itself was not worse in videos. Notably, the minimal effect size indicates no practical difference in retrieval between modalities ($\Delta R^2_{mar} = .003$), with a slight unexpected trend favoring videos.

The results suggest that videos were more effective than illustrated texts for learning science content for fifth- and sixth-grade students. The modality effects in cognitive load (Section 5.2), retention, and delayed retention are consistent with earlier findings for the modality effect in learning as well as predictions made by cognitive load theory (Sections 2.1–2.4). Figure 4 illustrates that the delayed retention was better in videos than illustrated texts in both research groups and in both climate change and balance in nature lessons. Furthermore, videos resulted in more consistent performance than illustrated texts in all learning metrics in both Study I and III. This can be seen concretely in Figure 4 for delayed retention. In other words, the learning outcomes are more equitable across students when learning from videos.

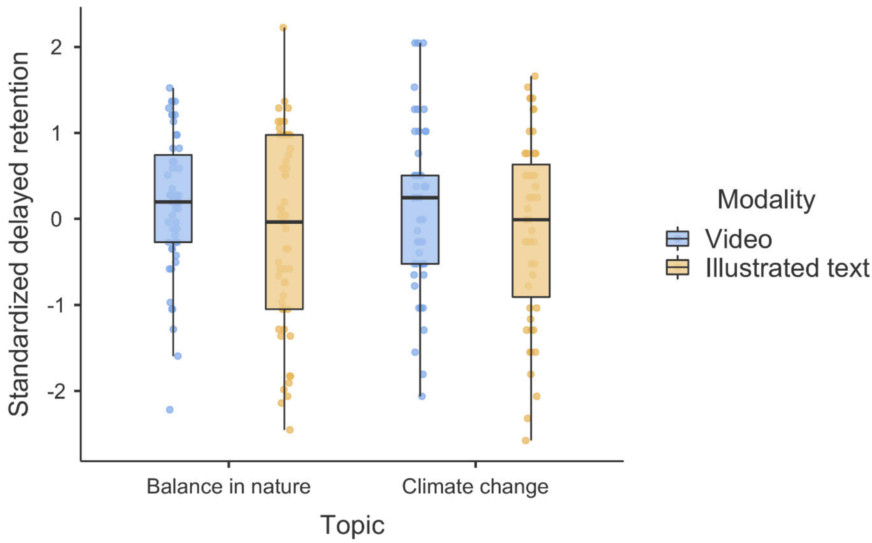

Figure 4. Standardized delayed retention across modalities and topics in Study III. Figure from Haavisto et al. (2025).

5.4 Modality effect in efficiency

Educators typically evaluate instructional methods not only by learning outcomes but also by the time and effort required, making efficiency an essential factor in decision-making.

In Study I, the videos led to lower cognitive load and better results across learning outcomes. An efficiency metric that takes into account both these benefits (Section 4.5) was higher for the videos compared to the illustrated texts in retention (R^2_{mar} = .107), delayed retention (R^2_{mar} = .139), and transfer (R^2_{mar} = .088), corresponding to a medium positive effect size in these outcome measures.

In Study III, the combined effect size of efficiency was calculated through combining the variances explained by the modality factor in both cognitive load and each learning metric, summing if the same modality led to both lower cognitive load and higher learning outcome and subtracting otherwise. Again, the video outperformed illustrated texts in efficiency for retrieval (R^2_{mar} = .049), delayed retention (R^2_{mar} = .062), and transfer (R^2_{mar} = .046), mostly driven by the lower cognitive load in videos. While the efficiency results were similar across studies, they were weaker in Study III. This is understandable, as the benefits of videos were weaker when material was available (Section 5.1), leading to more similar learning between modalities.

As learning time is the only factor that can be practically experimentally controlled, the present study chose to measure retention, retrieval, transfer, and cognitive load, while giving all participants the same limited time to study the learning materials. This also reflects a common situation in classroom teaching: students are often given limited time for individual studying before the class moves on to group activities or the next session. Learning time, learning outcomes, and cognitive load are all important for teachers, who wish their students to learn within the practical time constraints of a lesson but not at the expense of feeling exhausted at school. From this perspective, many previous studies of the modality effect among children are of limited use to teachers, because they often lack measurement of cognitive load or time limits.

Given these results, when taking into consideration learning time, learning outcomes, and cognitive load, the video was a more efficient way of learning in retention, delayed retention, transfer, and cognitive load.

Finally, previous research has noted that the advantages of learner control come mainly from increased time on task, as many studies have not controlled for total learning time (e.g., Li et al., 2022; Tabbers & de Koeijer, 2010). The present study shows that videos with hybrid learner control can improve learning outcomes even within realistic, time-constrained classroom settings.

5.5 The relationship between cognitive load and learning

Sections 5.2 and 5.3 showed that learning from videos was associated with both lower cognitive load and better learning in both Studies I and III. While both results align with the predictions of cognitive load theory, this does not necessarily mean that reduced cognitive load explains the improved learning. After all, the advantages of videos in learning and cognitive load could be two independent outcomes.

Raising further doubts about the significance of cognitive load explanations, Section 5.2 showed that the cognitive load levels were low or moderate in both studies and across conditions. Theoretical frameworks such as cognitive load theory and the cognitive theory of multimedia learning (Mayer, 2014, 2020, pp. 282–288) explain the modality effect primarily through the prevention of cognitive overload, rather than reducing cognitive load when learners still have cognitive capacity to spare. The findings suggest that cognitive overload may have been relatively rare in the primary school classroom context. Therefore, reducing extraneous load might not be a significant factor contributing to the advantage of videos in learning.

The predictive effect of cognitive load on each learning outcome was analyzed in Study I. According to mixed-effects analysis, cognitive load did not have an effect on retention, delayed retention, or transfer (p-values > .457). Furthermore, the correlations between cognitive load and learning outcomes were low in both illustrated texts and videos ($|r$-values$|$ < .081). Results of Study III further support this finding. The correlations between cognitive load and learning were low and similar between videos and illustrated texts in delayed retention (r = -.186 – -.192) and transfer (r = -.030 – -.043). Lastly, reading skills did not explain modality differences in cognitive load, contradicting the decoding explanation (Section 5.7).

These results suggest that the modality effect in learning was not explained by the reduced cognitive load, as described by prior research and prevalent theories. Instead, the advantages that videos provide in learning and cognitive load appear to be two separate effects.

Cognitive load theory was mainly developed in a laboratory research environment, where the learning situation is strictly controlled by the researcher, and the studying pace is often controlled by the system-paced learning material rather than the learner. Furthermore, in many studies, the materials and tasks were short, and the learner was expected to focus fully on just the task. In contrast, in primary education, learners typically control their pace themselves, and the tasks span from multiple minutes to a lesson, during which learners take spontaneous breaks, naturally avoiding cognitive overload (Section 2.10). Therefore, the traditional explanation of overload avoidance may not account for the modality effect in such settings, and the theoretical basis for the effect in learner-paced environments should be reconsidered.

Similar arguments were made by Schüler et al. (2011), who point out that the basic information-processing theories mostly focus on phenomena that occur over much shorter timeframes than multimedia learning in the classroom. As such, working memory effects might not provide the most useful level of explanation for learning in real-world educational scenarios where higher-level variables, such as motivation and cognitive or metacognitive strategies may hold more explanatory power (Schüler et al., 2011).

To conclude, several features of the present study may have contributed to the avoidance of cognitive overload in both modalities. Learners could progress at their own pace, they were given time to review the content, the familiar context did not require attention, and there was likely less pressure to perform than in a typical experimental environment. These aspects of classroom learning may help explain why the cognitive load levels were not related to the learning outcomes. The findings suggest that cognitive load theory should be applied with caution in realistic classroom settings, where learners are commonly not operating near their cognitive limits. Lastly, the findings of this dissertation suggest that the reduction in cognitive load and the improvement in learning outcomes associated with videos appear to be separate effects in a primary school classroom context.

5.6 The effect of individual differences on the modality effect

As cognitive load explained neither the learning results nor the difference in learning between illustrated texts and videos (Section 5.5), it is worthwhile to assess other factors. The most influential factor is individual differences, modeled with the participant-level random effects, which accounted for a major portion of the variability in all learning and cognitive load metrics in both Study I (R^2_{con} - R^2_{mar} = .436–.661) and Study III (R^2_{con} - R^2_{mar} = .462–.543). These results show that individual differences had a very large effect on participants' performance with both video and illustrated text materials. These effect sizes are much larger than the differences between the modalities. In other words, students who received high test scores with one modality often also received high scores with the other.

The high influence of individual differences suggests that they likely moderate the mechanisms behind the modality effect. First, the transient information effect depends on sustained attention and inhibitory control (Kim & Phillips, 2014), in which there are differences between individuals. Second, the additional cognitive load caused by decoding depends on individual differences in reading skills (Section 5.7). Third, there are individual differences in working memory capacity, which would affect all cognitive load explanations. Fourth, prior knowledge moderates many multimedia learning principles, commonly referred to as the expertise reversal effect (Kalyuga, 2007).

In contrast to the significant individual differences, the classroom-level random effects were excluded from all models because of the insignificant differences between classes (*intraclass correlations* = .013–.083). Furthermore, the school-level differences approached zero, and thus could not be estimated in many cases. This means that there is much greater variation between individuals than between classes or schools in how well students learn from multimedia in general, and how well they learn from videos compared to illustrated texts. This pattern is unsurprising, given that primary education in Finland is relatively unselective with respect to schools and classes.

While the models explained a large part of the variance in total, there is significant leftover variance that was not explained. This is due to unknown factors that are related to (a) how an individual changes between two consecutive lessons, (b) measurement error, or (c) the interaction between topic and modality (Section 7.1).

5.7 The effect of reading skills on the modality effect

As shown in Section 5.6, individual differences greatly influenced all outcome measures. Therefore, it is valuable to investigate the effect of specific individual differences. Study III examined the role of decoding ability by analyzing it as a fixed effect, separating it from other individual differences modeled through random effects. When controlling for modality, decoding predicted better retrieval from materials, transfer, and retention with medium to large effect sizes (ΔR^2_{mar} = .169, .086, and .098, respectively). This aligns with expectations: decoding indirectly predicts better performance across modalities, because it correlates with broader verbal ability (Section 2.5).

However, the increase in total variance explained was minimal (ΔR^2_{con} -values < .011), suggesting that variance was reassigned from random effects to decoding. This demonstrates that the initial strategy of accounting for individual differences through random effects was successful with respect to decoding.

The interaction effect was analyzed by adding it to a model with a fixed effect for modality and decoding. The interaction between decoding and modality predicted retention with a small effect size (ΔR^2_{mar} = .013): videos were particularly beneficial for children with lower decoding abilities.

Figure 5 shows the magnitude of this interaction. Videos led to significantly better delayed retention for children in the weakest decoding ability quartile. Conversely, the strongest decoding quartile had only small differences between modality conditions. Although other interactions were not significant, videos also descriptively especially benefited those with weak decoding in transfer (p = .071).

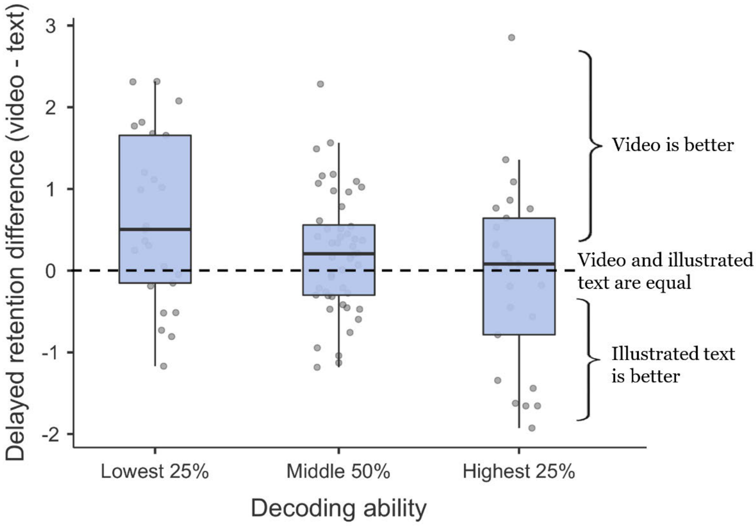

Figure 5. Standardized delayed retention across modalities and topics. Figure from Haavisto et al. (2025).

Surprisingly, decoding did not predict cognitive load when accounting for modality. Furthermore, the interaction between decoding and modality did not predict cognitive load. This suggests that decoding skill does not meaningfully explain the differences in cognitive load when learning from illustrated texts versus videos. This finding contradicts the explanation based on cognitive load theory, which predicts learners with better decoding ability would experience less effort when learning from narrated multimedia (Section 2.4). Since the learners were children, decoding ability was expected to be even more influential than in prior studies, which focused mostly on university students. Instead, the results suggest that fifth and sixth graders had sufficient decoding ability to avoid experiencing decoding as particularly effortful in this primary education context, repeating the pattern of results discussed in Section 5.5.

Similarly to decoding, reading comprehension did not directly predict cognitive load nor did it interact with modality to predict cognitive load differences between modalities. The learning results also followed the same patterns as decoding. After accounting for modality, reading comprehension improved retrieval from materials, transfer, and retention with medium to large effect sizes (ΔR^2_{mar} = .222, .115, and .276, respectively; p-values < .004). Again, video especially benefited children with weak reading comprehension in delayed retention (ΔR^2_{mar} = .009, p = .036) but other interactions were not significant.

Compared to decoding, the main effect of reading comprehension was much greater in effect size, but the interactions with modality were weaker. This result aligns with the simple view of reading (Section 2.5), in which decoding is conceptualized as a component within reading comprehension. Consequently, reading comprehension offers a more general benefit for studying both modalities, whereas decoding more narrowly supports learning from illustrated texts.

Notably, when decoding and its interaction with modality were included in the analysis, the main effect of modality remained significant. This suggests that videos offer advantages beyond simply bypassing the cognitive process of decoding written words within working memory. This is particularly relevant for cognitive load outcomes, as decoding was not directly related to cognitive load, nor did it interact with modality to predict cognitive load. At the same time, modality showed the strongest influence on cognitive load among all the outcome variables. This finding along with findings from Section 5.5 suggest that the conventional explanations for the modality effect (Section 2.4) may need to be reconsidered in the context of primary education.

5.8 Modality effect in studying rate

Although children's reading abilities explained some of the learning benefits of videos, the reduced cognitive load seen in the video condition appears to arise from other advantages than avoiding the need to decode written text. One potential explanation is related to differences in studying rate between modalities, which depend not only on the learner's reading skills, but also on modality-specific learner controls, such as the automatic progression feature of videos. After all, the use of available navigation options (e.g., pausing in the video or pressing a button to go to the next slide in the illustrated text) directly shapes how quickly children progress in the materials.

Study II explored how fifth-grade students study videos and illustrated texts and how these study behaviors relate to learning outcomes and cognitive load. A total of 97 screen recordings from the learning phase in Study I were analyzed, each displaying the learning process of one fifth grader. The recordings were coded frame by frame to obtain each child's time spent on each slide, the order of accessing slides, and the used navigation actions. The time and order data were further analyzed and categorized into studying rates and strategies (Section 4.7).

A mixed-effects analysis revealed that modality affected the studying rate with a strong effect size ($p < .001$, $\Delta R^2_{mar} = .148$). The children studied the videos 37%, 20%, and 3% faster at the 25th, 50th, and 75th percentiles than they studied the illustrated texts. Considering the modality effect (Sections 5.2, 5.3) and the effect of reading skills (Section 5.7), this suggests that the videos offer direct benefits especially for those who read more slowly.

Words per minute (wpm) and pictures per minute (ppm) studying rate in the videos was 112 wpm + 3.2 ppm, 119 wpm + 3.4 ppm, and 120 wpm + 3.4 ppm for the 25th, 50th, and 75th percentiles, respectively. The studying rate in the illustrated text condition was slower: 82 wpm + 2.3 ppm, 99 wpm + 2.8 ppm, and 117 wpm +

3.3 ppm for the 25th, 50th, and 75th percentiles. For reference, the videos were presented at 121 wpm + 3.4 ppm.

Figure 6 illustrates the faster and more consistent studying rates in the initial pass of videos compared to illustrated texts. The mostly parallel lines in the upper spaghetti plot demonstrate that most learners watched the videos at their designed pace, and thus the studying rates were highly similar. In contrast, the studying rates when learning from illustrated texts were highly varied. Despite this variance, the backward button was still used only 1.5 times on average during the initial pass of the illustrated text. This means that the variation in reading duration was largely not due to returning to earlier slides, but due to variations in reading rate.

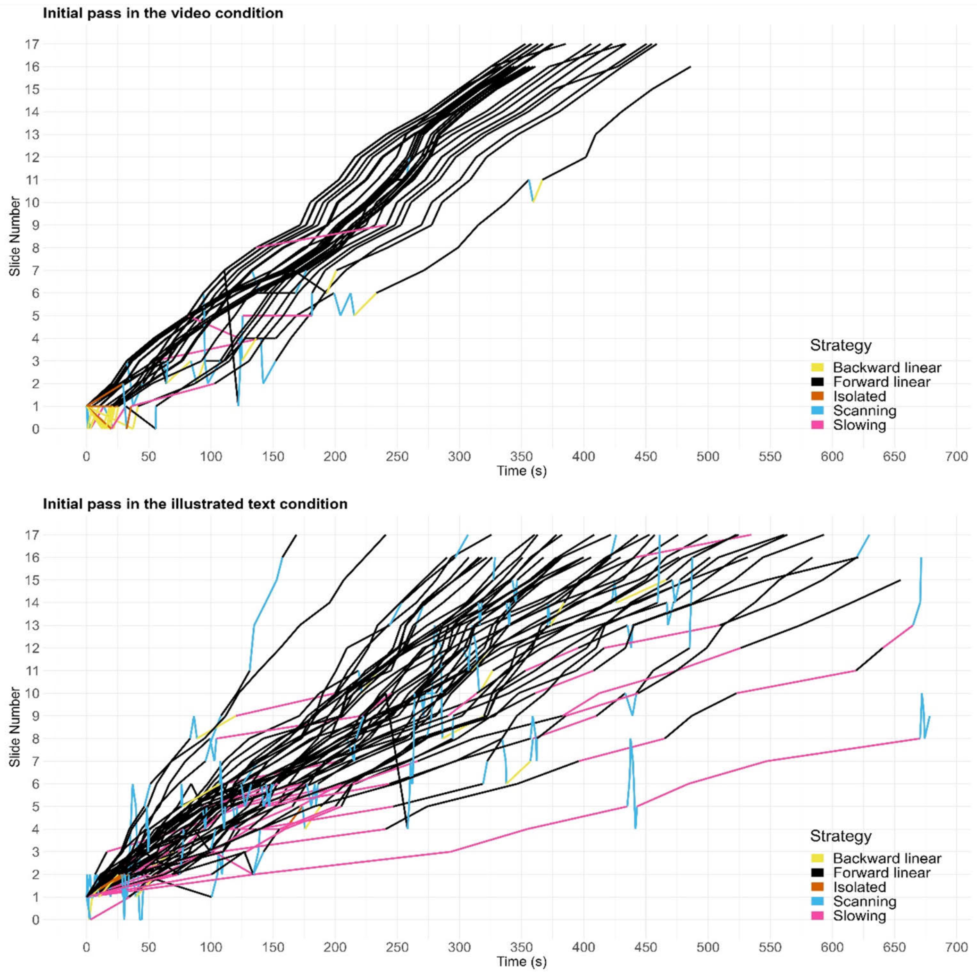

Figure 6. Each child's progression through the slides during the initial pass of videos and illustrated texts. Each child studied in both conditions. Slide number 0 denotes the title slide. Figure from Haavisto and Jaakkola (2025).

A surprising finding, considering the high influence of individual differences (Sections 5.6 and 5.7), is that the effect of individual differences on studying rate was weak (individual-level random effects: R^2_{con} - R^2_{mar} = .011). This indicates that the studying rates of the same child were almost completely unrelated across modalities. A likely explanation is that videos encourage steady progress through instructional content because of automatic progression, as the video studying rates varied little (further investigation in Section 5.9). In contrast, reading rates differ significantly between individuals, which may be one reason why the illustrated text condition descriptively showed larger variance in outcome measures in Studies I and III compared to the video condition (Section 5.1).

The relatively quick studying rate observed in the video condition can be explained by the children's limited use of navigation. This finding is consistent with earlier research showing that learners with little prior knowledge seldom make use of navigation features. Furthermore, maintaining a steady, linear approach to studying was beneficial. This allowed the children to avoid a common drawback of learner control: novices might overlook essential information or disrupt the logical flow of explanations through poorly chosen navigation paths. Notably, the transient information in the videos did not prompt learners to pause or replay parts of the video, which may reflect the relatively low levels of cognitive load when studying longer materials in an authentic learning environment with less pressure to perform compared to a laboratory context (Section 2.10).

Because the studying rate in the videos closely matched the narration speed of the videos, comparing studying rates with ones reported in previous studies may offer limited insight. Still, the rate of narration in the videos used in this dissertation (121 wpm) was slower than educational videos on YouTube (145 wpm; ten Hove & van der Meij, 2015), edX courses (156 wpm; Guo et al., 2014), and the most efficient estimated listening speeds for children (228–328 wpm; Woodcock & Clark, 1968). Therefore, faster-paced videos than those used in this dissertation could further enhance learning efficiency, especially since transient information was not a significant issue.

Also, the studying rate in the illustrated texts is difficult to compare with earlier research. After all, no previous studies match all the key factors that influence studying rate: the learners' young age and reading skills, studying for recall performance, the use of pictures in the learning materials, and the classroom context. Even so, studying rates observed in Study II are considerably slower than the typical on-screen reading speeds reported for fifth graders, which range from 142 to 207 words per minute (Spichtig et al., 2022). However, these faster rates are obtained using a comprehension-focused silent reading paradigm, where children read shorter, isolated texts followed by comprehension questions, and where the data are cleaned to remove unusually fast or slow progressions. Therefore, the figures reported in earlier research may not represent typical studying rates in classroom contexts.

The slower studying rate in the illustrated texts suggests that children take longer when processing information presented in written multimedia. This may stem from both their limited reading skills and the need to divide attention between the written text and the accompanying images. A contrary explanation for the slow reading is that the students engaged in frequent rereading within slides to actively connect ideas. However, this seems unlikely, as active rereading has been linked to improved learning outcomes (Section 2.8), while the illustrated texts resulted in lower performance compared to the videos, in which within-slide rewinds were infrequent.

Another potential reason behind the slow reading is that the children's attention shifted to something other than the material, especially given the lower learning outcomes, and the fact that the illustrated texts did not progress unless the reader used navigation options. This meant that accessing additional content in the illustrated texts required effort, motivation, and self-regulation, while the videos advanced through the content automatically.

The hybrid learner control may help explain why the videos led to better outcomes than the illustrated texts, as meta-analyses have found no benefit in studies where learner-controlled narrations did not progress automatically (Ginns, 2005; Reinwein, 2012). However, there are also other differences that may explain the contrasting results. First, earlier studies have generally involved university-level participants, who read much faster than primary school children, whereas there is less difference between adults' and children's listening rates. Second, the participants were encouraged to use the remaining time for reviewing. Therefore, benefits in studying rate increased the time for reviewing, which supported learning outcomes rather than decreasing studying time. This highlights the role of instruction in learner-controlled contexts. Indeed, in some prior studies, even though participants could study for as long as they wished, they commonly spent roughly only the time required to study the materials once (Section 2.7). Third, the use of YouTube videos in this dissertation likely provided a more familiar format for children than the slideshow presentations with embedded audio that were common in earlier studies. This familiarity may have reduced negative effects of unfamiliar interfaces, which can initially interfere with learning (Section 2.7). Relatedly, the hybrid learner control may have minimized the cognitive load associated with managing the presentation thanks to automatic progression.

5.9 The effect of slide characteristics on studying rate

Despite modality and individual differences explaining studying rate to a significant degree, 84.1% of the variance in the studying rate remained unexplained. Thus, studying rate was analyzed more closely at an individual slide level. As the mechanism that explains the studying rate was expected to differ between the modalities, this analysis was done separately for illustrated text and video data. Stepwise regression was used, and fixed effects were three slide characteristics (number of text characters, slide number, and picture type) that were added into the models in order from greatest to least variance explained.

The number of text characters explained half of the variance in the initial pass slide durations in the illustrated texts (ΔR^2 = .500, p < .001) and most of the variance in the videos (ΔR^2 = .898, p < .001). In other words, the text quantity accounted for most of the variation in how long slides were studied across both formats. This was especially true for videos. Learners did not use the videos' learner controls much, and therefore the videos progressed along with the narration rate, which in turn was almost completely determined by text content. In Illustrated texts, it also makes sense that an increased number of words takes a proportionally longer time to read.

The second most significant predictor for the studying rate in illustrated texts was the slide number, which affected study durations with a large effect size (ΔR^2 = .284, p < .001). In the videos, the number of the slide was also the second most significant predictor, with a small effect (ΔR^2 = .039, p < .001). This phenomenon of earlier slides using more studying time was about seven times stronger in the illustrated text condition, which reflects the automatic progression of the videos: there was less leftover variance to explain as the amount of text already explained most of the variance in video studying rate. The dedication of more time to the initial slides of the illustrated texts aligns with earlier findings showing increased fixation durations (Hyönä & Niemi, 1990) and slower reading (Hayden et al., 2019; Hiebert & Daniel, 2019) at the beginning of texts. One possible explanation is that students reread less over time as they become tired, lose motivation, or become concerned about time constraints. Another possibility is that they increase their reading speed as they become more familiar with the task. In any case, the considerable differences in how long students spent on early versus later slides in the illustrated texts indicate that the available time was not used consistently. This uneven distribution may partly explain their weaker performance on the retention tests, which included questions covering the entire content.

The picture type (photograph or diagram) did not significantly affect the initial pass slide duration in the illustrated texts (ΔR^2 = .020, p > .05) or the videos (ΔR^2 < .001, p > .05). This minimal effect of picture type across modalities contrasts with the great emphasis given to the role of images in multimedia learning research, especially since this dissertation included relevant pictures (Section 4.4). While the results are again similar across modalities, they probably have different causes.

In the videos, the result suggests that the type of image contributed minimally to the variation in time spent on each slide. Very little variance (6%) was left to be explained after the number of characters and slide number had been controlled, which again highlights the hybrid learner control in videos. This implies that even the inclusion of images itself may have had limited impact on study time. Indeed, previous studies have found that the most efficient studying rates are similar in sped-up narrations, in both multimedia materials (Pastore, 2012) and materials without pictures (Heiman et al., 1986).

In the illustrated texts, the study pace was not influenced by the image type, which is surprising given that illustrated texts do not progress automatically and learners cannot attend to both text and images at the same time. However, two factors may potentially reduce the effect of picture type. First, students may have spent only a limited amount of time examining the pictures, which corresponds with some earlier results (Hannus & Hyönä, 1999) while contrasting with others (Mason et al., 2013). Second, the cognitive load was low to moderate, which may have allowed readers to invest extra effort in processing the images without exceeding their cognitive capacity (Park et al., 2011, 2015).

In conclusion, in both modalities, just the text content and the slide number determined the studying rate to a great degree. However, the slide-level factors indeed differed between modalities: video studying rate was overwhelmingly determined by the narration rate, whereas illustrated text studying rate was also largely affected by the slide position.

5.10 Modality effect in review strategies

Figure 7 illustrates that the review strategies (Section 4.7) of participants differed significantly between illustrated texts and videos. As can be seen from the spaghetti plots, most (~80%) of the video reviewing was forward linear, whereas in the illustrated texts, no review strategy was as prominent. The prevalence of all review strategies, except for the isolated strategy, differed between videos and illustrated texts with a medium to large effect size (ΔR^2_{mar} = .139–.409). In contrast, individual differences modeled with random effects were less significant, which means that the review strategies used by the same individual varied between modalities.

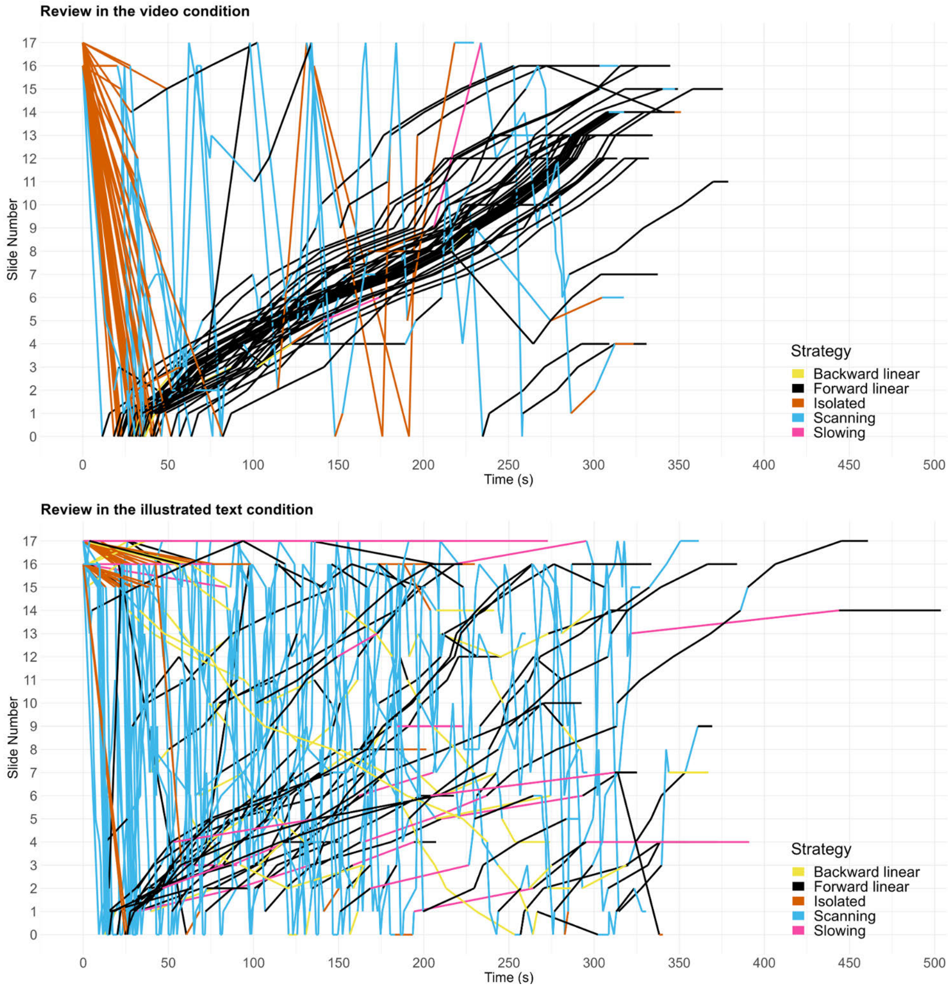

Figure 7. Progression through the slides during the review of videos and illustrated texts. Each child studied in both conditions. Slide number 0 denotes the title slide. Figure from Haavisto and Jaakkola (2025).

The videos were generally reviewed starting from the beginning, with only a small portion of the review time involving navigation actions. Most of the time was spent watching the videos at their intended pace and sequence, as was the case during initial studying (Section 5.8). This pattern supports the idea that the hybrid learner control in videos shapes learners' strategic behavior. The mostly linear behavior in both initial study and review means that most children attempted to watch the videos twice, although learning time ran out before most finished the second playthrough. Given the strong learning outcomes associated with the videos, these results are consistent with earlier findings that watching videos twice is a good strategy for retention, perhaps because rewatching allows learners to focus on integrating information (Section 2.8).

In contrast, when reviewing illustrated texts, children showed more varied behaviors, including linear backward, scanning, and slowing. This is consistent with the idea that children often apply multiple strategies (Section 2.8). Some children either reviewed systematically from the end toward the beginning or stayed on the same slide without advancing. These patterns reflect how the lack of automatic progression shapes learners' studying rate and order. In the illustrated texts, taking no action resulted in slowing, and progressing backward was just as easy as forward. In contrast, slowing in the video condition required learners to actively press the pause button, and reviewing in reverse order was almost impossible, requiring learners to precisely rewind to each earlier slide after finishing the subsequent one.

Finally, the slower studying rate in the illustrated texts during the initial study, compared to videos, may have contributed to the use of different review strategies due to increased pressure from limited time. After reading through the material once, few learners would have had the time to read it for a second time from start to finish.

5.11 Synthesis of main findings

The modality effects were clearly present across studies. Videos yielded better retention and delayed retention, lower cognitive load, and similar transfer and retrieval outcomes (Sections 5.2–5.3). Because both modalities were studied for equal durations, videos were more efficient when considering learning outcomes, cognitive load, and learning time (Section 5.4). Individual differences were substantial across all learning outcome and cognitive load measures (Section 5.6), and videos were particularly beneficial for learners with weaker decoding skills in delayed retention (Section 5.7).

There were also clear differences in studying behavior between conditions. Videos led to faster and more uniform studying rates, with learners largely following the progression of the video, whereas illustrated texts resulted in slower, more variable rates and diverse review strategies (Sections 5.8, 5.10). These findings are novel, as they were demonstrated in the primary school classroom contexts using typical multimedia learning materials (Section 2.10).

In this context, the modality effects could not be explained by differences in cognitive load. Cognitive load levels were mostly low to moderate, and cognitive load did not explain the learning differences between modalities (Sections 5.2 and 5.5). Furthermore, learners with better decoding ability did not experience significantly less effort when studying illustrated texts (Section 5.7). This is especially surprising, as the participants were primary school children with developing reading skills, and decoding is directly related to a key source of cognitive load: converting written text into verbal information within working memory (Section 2.4). These findings challenge the common explanation that the modality effect is based on preventing cognitive overload (Section 2.4).

In study II, the relationship between modality, learning process, and outcomes was further analyzed by examining how studying rate and review strategies predicted learning and cognitive load measures. Without controlling for modality, faster studying predicted higher retention, delayed retention, and transfer as well as lower cognitive load (ΔR^2_{mar} = .032–.081). After controlling for modality as a fixed effect, faster studying still predicted better transfer (ΔR^2_{mar} = .054) but was no longer significant for other outcomes. This aligns with videos being studied faster and outperforming illustrated texts in outcomes other than transfer (Sections 5.2, 5.3, and 5.8). A similar pattern occurred for review strategies. A higher proportion of forward linear strategy (more common in the video condition) predicted better retention and delayed retention (ΔR^2_{mar} = .017–.052) and lower cognitive load (ΔR^2_{mar} = .180). When modality was controlled for, the effect on cognitive load decreased (ΔR^2_{mar} = .095) and the effects on learning outcomes disappeared. These results suggest that the predictive effects of studying rate and strategy on learning outcomes and cognitive load were mostly attributable to differences between modalities. In other words, faster and more linear studying can be seen as mechanisms through which modality influences retention and cognitive load.

Together, the findings in this dissertation support a straightforward interpretation: most children read more slowly than the video narration progressed, and the videos effectively standardized the learning pace. This resulted in less variance in learning outcomes across learners (Section 5.1, Table 4), and especially benefited slower readers (Section 5.7). In practical terms, faster and more linear studying was likely advantageous because (a) covering the material quickly during the initial pass allowed more time for review, and (b) studying in order from start to finish distributed the limited learning time evenly across the material. Slide-level analyses (Section 5.9) confirmed this: in the illustrated text condition, children dedicated substantially more time to earlier slides, resulting in a much less balanced allocation of time compared to videos.

In summary, the present dissertation observed the modality effect, as predicted by cognitive load theory. However, the context of the present study differs substantially from previous studies in both materials and settings. Cognitive load theory originated in contexts where individuals typically perform at peak capacity for short periods in controlled environments. In the context of fifth and sixth graders learning from illustrated texts and videos in the classroom at their own pace, learning differences between modalities often cannot be explained by the avoidance of cognitive overload. In such contexts, the explanations for the modality effect may stem less from lower-level working memory effects, and more from higher-level, context-dependent processes that unfold over longer timescales.

6 Methodological Implications

6.1 Scientific rigor in an uncontrolled environment

When one opens the methodology section of a research paper in educational psychology, they often find descriptions of a lab environment with strict control. Then, opening the discussion section would commonly show recommendations on how the results should be taken into account in the classroom. However, the testing of effects outside of the classroom is likely to produce results that do not generalize to the classroom (Section 2.10). Crucially, researchers have the burden to prove why each difference between the lab and classroom does not matter. This burden is heavy, as a change in the environment is likely linked to changes in outcomes through a complex chain of events (Higgins et al., 2005).

The usually cited reason for studying learning in a lab environment is scientific control. Tightly controlled lab research is valuable for isolating specific variables and for examining micro-level processes that are hard to observe in classrooms. However, such control becomes undesirable when studying learning as it unfolds in real settings because it affects the very phenomenon under investigation. A concrete example is someone knocking on the classroom door. Such disruptions may affect students differently, depending on whether they are learning from videos or illustrated texts. The knocking might be more disruptive to those listening because it could be harder for students to find the spot in the narration where they got distracted. On the other hand, listening might be less affected if the narration makes the students not even hear the interruption. In either case, removing disturbances introduces systematic error by masking inherent advantages of listening and reading. Studies that test experimental conditions separately must control the environment in such unnatural ways to prevent confounding factors. In contrast, when both experimental conditions are implemented simultaneously in the same environment, the researcher does not have to control (or even be aware of) the specific phenomena that occurred (e.g., disruptions).

In addition to avoiding systematic bias, environmental variation also enhances the generalizability of the results to a broader range of settings. For example, instructions on how to study the materials during interventions can be explained in a natural way, varying across each studied class. As a result, the findings are more robust to differences in instructional delivery and not dependent on a particular scripted instruction. After all, teachers will not adhere to a script when they implement instructional methods in their classroom.

While the variation resulting from a less controlled environment (e.g., students sitting in different spots in the classroom) could reduce statistical power, this turned out to not be an issue in this research, since the proportion of variance explained by the mixed-effect models was high.

This study applied a similar approach to individual differences as it did to the classroom context, aiming to account for all factors as a whole without isolating specific ones. The findings indicate that individual differences greatly influence the modality effect, explaining much of the variation in outcomes across all tests. In addition to relatively stable factors like age, which are often accounted for in research, the within-subjects design allowed the study to also take into account more variable factors. These include temporary states of individuals that can shift daily, such as the quality of sleep the night before or how a participant interprets a subjective cognitive load question. Since it is not realistic to measure every individual or contextual factor that might influence learning and cognitive load, within-subjects designs in authentic settings are particularly well suited for investigating learning effects.

6.2 Ecologically valid manipulation of the modality

The above section showed how the ideal of scientific control in the environment can hinder generalizability or even confound the results, while this section discusses related problems in the manipulation of the modality variable.

Studies of the modality effect aim to isolate the modality variable by using identical materials except for replacing written text with narration or vice versa. This appears simple in theory but is complex in practice. Converting materials between modalities involves several design choices that are not equivalent between modalities. Written materials involve text layout and font, while narrations have different tones of voice and playback options. Replacing the text with narration requires decisions about who speaks and at what speed. Such decisions are often not documented and they could potentially influence the modality effect.

Since learning materials used in classrooms differ in multiple ways but studies attempt to change just one aspect, modality research has often ended up studying materials that would rarely be used outside a research context. To make this more concrete, consider a common form of illustrated text: a PowerPoint presentation with text on the left and an image on the right. If the text is replaced with a narration, the left side of the presentation is blank. While this may seem like a rigorous manipulation of the modality, it actually diminishes generalizability because half-blank presentations are not used in schools. This leads to an unfair modality comparison because it does not reflect inherent strengths of narrated multimedia, such as having more space for pictures.

Similar problems would arise from taking a video as a starting point and minimally altering it to arrive at a written modality condition. This could be done for example by muting a video and turning on subtitles. While the resulting material indeed only differs in that the narration is replaced by written text, the result does not utilize the strengths of written text. First, the learner cannot adjust the pace of progression to their reading speed. Second, the learner cannot fluently go back and forth in the text by moving their gaze. Third, if the material has dynamic visualizations, the learner would miss visual content while they are reading the subtitles. These reasons explain why muted videos with subtitles are not a commonly used type of illustrated text in education.

A related problem arises also at the research procedure level. In some studies of the modality effect, the time available for learning in both conditions was set by the length of the narration (Tabbers et al., 2001). This does not make sense for the written condition, as reading speeds vary widely. For example, in Study II, the fastest reader read the text several times faster than the slowest one, whereas most students took approximately the same time to watch the video. Imposing a fixed reading time based on the length of the narration also does not reflect classroom practice, as teachers do not allocate reading time based on how long it would take to listen to a corresponding narration.

To conclude, the modalities inherently differ in several ways. Therefore, efforts to change only one variable for purposes of scientific purity lead to results that (a) could be biased toward the modality that was chosen as a starting point and (b) might not generalize to classroom implementations. This dissertation addressed the methodological challenge by designing both illustrated text and video conditions as authentic examples of how each modality is used in classrooms, allowing each material to leverage its advantages. The students were given the same time to study both modalities, which was nearly double the narration length in Study I and II, and almost six times in Study III. This fairly represented learning from both materials in a time-limited lesson context.

6.3 Efficiency

An important methodological consideration guiding this dissertation was evaluating the efficiency of learning by jointly taking into account the learning outcomes, cognitive load, and learning time. Learning and cognitive load were measured while learning time was limited and equal across modality conditions. This approach was due to the fundamental differences between these three variables: unlike outcomes and cognitive load, learning time can be set by the researcher.

Previous studies have often lacked this consideration (Section 2.9). In many cases, learning time was not limited or measured, leaving open the possibility that the learning advantages of a modality were simply due to more time spent learning. Furthermore, a combined efficiency metric has often not been considered. This is problematic when the results are better in one aspect but worse in another. For example, enabling learner control commonly increases time on task. In such cases, learner control might not be recommended even if it leads to better learning. Similar considerations apply when learning is enhanced with the cost of increased cognitive load. It could be argued that the recommendations of a study are open to doubt if even one of the three is missing. After all, instructional design aims not just to maximize learning, but to do so in a way that fits within practical time and effort constraints.

Two approaches are recommended to improve the assessment of classroom effectiveness and the comparability of results across studies. First, as done in this dissertation, studies with limited learning time should report both learning outcomes and cognitive load, and match the learning time between instructional conditions. Second, studies with unrestricted learning time could report a new metric that incorporates learning outcomes, cognitive load, and learning time, extending existing concepts such as relative condition efficiency (Paas & Van Merriënboer, 1993) and time-efficiency (Rasch & Schnotz, 2009). One reasonable example of such a metric, which weighs each factor equally, would be to divide the standardized learning outcome by the product of standardized cognitive load and standardized learning time. Of course, the weights assigned to each factor depend on the specific goals of the learning. Moreover, while benefits in improved learning and reduced time might be considered roughly linear, the value of lowering cognitive load may be nonlinear, with the greatest benefit occurring when it prevents cognitive overload.

However, these three factors do not capture all relevant aspects of instructional quality; a fourth factor such as enjoyment would be a valuable addition. If learners willingly spend more time and cognitive effort with a particular modality, this may indicate higher motivation. In that case, the greater investment of resources could be interpreted as an inherent strength of the modality.

Lastly, there is a deeper assumption behind these efficiency metrics, one that learning with less time and effort is preferable. But this is not necessarily true. If students are enjoying the activity, longer learning time might be desirable. Furthermore, in this dissertation, cognitive load was at a moderate level and did not negatively affect learning. In this situation, extra cognitive load might not result in any overload or exhaustion.

6.4 Interpreting nonsignificant findings in modality research

Several studies of the modality effect have concluded that there is no effect simply because the *p*-value exceeded .05. However, this interpretation misrepresents the logic of null hypothesis testing. A nonsignificant result does not imply the absence of an effect; it only indicates that the data do not provide strong enough evidence to reject the null hypothesis. Drawing conclusions based on such findings is incorrect when the study lacks adequate statistical power to detect effects of meaningful size. Indeed, many studies on the modality effect suffer from small sample sizes and underpowered designs. For example, the meta-analysis by Reinwein (2012) found that the average sample size per condition in between-subjects designs was about 19 participants. When comparing two conditions, such as narrated versus written modalities, a sample of this size only allows for detecting large effects ($d > 0.93$) with sufficient power (80%) at the conventional significance level ($p < .05$; Faul et al., 2007). This means that many medium or small effects, such as those reported in this dissertation, may have gone undetected and misinterpreted as evidence of no effect.

A common pattern in prior research is to attribute the absence of a modality effect to a specific moderator variable. That is, when no effect is observed, the result is interpreted not as a lack of statistical power but as support for the influence of a moderator. However, this line of reasoning is only valid if the study has sufficient sensitivity to detect an effect in the first place.

The present study addressed these concerns by employing a design in which each participant served as their own control. This approach substantially increased statistical power by accounting for most of the variance across most tests. In addition, a method for approximating statistical power was developed in Study III. The method simulated a thousand bootstrapped models per iteration, and iteratively adjusted the target effect size in increasingly smaller steps, decreasing it when power exceeded 80% and increasing it when below, until identifying the effect size corresponding to 80% power. This method can be applied in any study regardless of the statistical tests used to approximate post hoc statistical power.

To further improve the interpretability of nonsignificant findings, equivalence testing could be used (Lakens et al., 2018). For example, Tarchi et al. (2021) illustrate how equivalence testing can clarify whether a nonsignificant modality effect truly suggests no meaningful difference.

7 Limitations and Suggestions for Future Research

7.1 Study design

One drawback of the study design in this dissertation is that the interaction effect between the order of modalities and the topic is unknowable. For example, videos might be particularly beneficial for certain topics depending on whether they follow or precede illustrated texts. However, in classrooms, materials are often used in various orders, reducing the likelihood of this being a serious confounder. Moreover, the topics were designed to be qualitatively similar in many ways: difficulty, length, style, and declarative knowledge, which reduces topic effects. As individual differences explained the largest portion of the variance in most variables, the benefits of the chosen study design outweighed the potential influence of interactions between orders and topics. The study design is therefore recommended in future studies, as it greatly improves explanatory and statistical power.

In contrast to the large individual differences, the variance explained by the modality conditions was quite modest across all measures, and the differences in transfer were not practically significant. It is understandable that a relatively short two lesson intervention was not able to make a difference in transferable understanding on topics as broad as climate change and balance in nature. However, another perspective is that the benefits of videos were already observable in many measures after a relatively short intervention. Future longer-term interventions could determine whether the benefits accumulate over time with continued studying from videos compared to illustrated texts. Lastly, a relatively short intervention could lead to novelty effects. However, this dissertation emphasized ecological validity, which reduces novelty (Section 2.10).

7.2 Measures

One limitation in measures is that the cognitive load self-assessed after studying is different from the level of cognitive load experienced at each moment during studying. The cognitive load explanations discussed in this dissertation depend on the assumption that the measured higher average cognitive load would mean that there were more instances of cognitive overload. This limitation was accepted because there are no reliable real-time measures of cognitive load that would preserve the authentic classroom environment. Furthermore, research suggests that real-time measures of cognitive load (e.g., pupil diameter, skin conductance, heart rate) are less reliable than self-assessment (Section 4.8.4).

Likewise, eye-tracking was not used because it would have disrupted the authentic classroom environment (Section 2.10.3). Therefore, this research was unable to determine where the students were looking while studying: the pictures, the text, or away from the material. Furthermore, the screen recordings only allowed tracking of within-slide progression in the videos, whereas in the illustrated texts, it only showed which slide was visible. In future research, non-invasive types of eye-tracking hold great potential, for example, using built-in cameras of students' own tablet computers.

Screen recordings were not assessed in Study III, and therefore, the studying process during the retrieval from materials tasks is unknown. The availability of materials during testing likely encouraged more active navigation, as the retrieval questions required locating specific information. Future research is suggested to examine learners' navigational behavior in contexts where materials remain available during testing.

While this study investigated the roles of cognitive load, reading skills, and studying process in determining learning outcomes in videos and illustrated texts, a large part of the variance was unexplained. The results suggest that variables not measured in this dissertation play a significant role in the modality effect. Particularly, future research is recommended to assess the effect of metacognitive and motivational variables on the modality effect.

7.3 Scope

Recommending video-based learning for those with weaker reading skills can be problematic since the children would get even less reading practice, which would not help them improve their reading skills. However, the primary focus of this dissertation was on supporting science learning, which is especially relevant considering recent PISA findings, which show simultaneous declines in both reading literacy and science performance. As such, it is important to identify effective instructional strategies that support content learning even when reading skills are in global decline. Second, skills such as reading, listening, or video-studying skills are typically addressed through dedicated instruction, and the goal of science lessons is usually to promote understanding of scientific content. Nevertheless, future research is suggested to measure how learning from different modalities improves the learning skills in that specific modality.

This dissertation studied just nine classes and four schools, so the question arises whether it generalizes to other classes and schools. As individual differences vastly outweighed class- and school-level differences, the results likely generalize to similar classes. However, all of the classes share characteristics of the broader Finnish education system. Therefore, future research is needed to examine similar interventions in other educational environments to validate the results.

This dissertation only addressed individual learning. The children learned from studying multimedia materials, and by answering test questions with or without the material available. Also, the research was limited to two science topics. However, the results were similar across these study methods and topics, suggesting potential generalization to other studying methods and science topics.

While the dissertation tried to match common materials used in classrooms today, they might not represent different types of illustrated texts and videos. For example, this study used a specific font in the written text and a specific speed of the narration. Particularly testing faster video speeds has potential for future research as the default narration tempo mostly determined the studying rate and the videos were relatively slow and cognitive load only moderate. Thus, faster videos may benefit learning even more.

7.4 Ecological validity

Although the study was designed to closely replicate classroom learning, the children knew they were part of a research project due to ethical requirements. This awareness may have influenced how they completed learning tasks. The study setting was such that the researcher was a visiting co-teacher, which might make students react differently compared to a lesson instructed only by their own teacher. As discussed in Section 6.2, it is possible that these characteristics affected one modality more than another. However, the teacher effects might be lesser as content was learned individually, and the role of the teacher was just to distribute the materials and allocate lesson time.

Some aspects of ecological validity could not be included in the study to preserve the comparability of conditions. One such limitation is that the materials only used static visualizations. Although dynamic visualizations are ubiquitous in videos, they are not common in illustrated texts. Therefore, their inclusion in the illustrated text condition would have compromised the ecological validity of the experiment. On the other hand, including dynamic visualizations only to the video condition would make comparisons between the conditions arbitrary, since dynamic and static visualizations inevitably differ in content. Similarly, to preserve content-equivalency, the fact that videos usually have more pictures per word compared to illustrated texts could not be reflected in the learning conditions. As the materials studied in this dissertation did not fully utilize these features, videos may hold even greater potential.

This study aimed to represent the authentic classroom context as a whole and did not try to identify specific contextual factors by manipulating them. This is because ecological validity cannot be represented by the sum of its parts (Section 6.2). Therefore, future studies are cautioned against manipulating specific components of ecological validity individually.

8 Practical Implications and Conclusion

Despite the widespread use of videos in primary education, their relative advantages over illustrated texts have been uncertain, largely because few studies have examined how these materials are used in the primary school classroom context. The goal of this dissertation was to address gaps in previous multimedia learning research, which has been limited in terms of (a) ecological validity, (b) testing the role of cognitive load and reading skills in influencing the modality effect, and (c) analyzing differences in the learning processes between modalities.

The findings of the present dissertation suggest that videos are an effective tool for science learning in primary school. Learning from videos led to better retention, lower cognitive load, and higher instructional efficiency than learning from illustrated texts. Because the learning time was equal across modalities, videos were also more time-efficient. Such efficiency is crucial for classroom learning, where time is limited and high cognitive load during learning is not a sustainable strategy. Videos were also studied faster than illustrated texts, and they were equally effective in supporting both information retrieval and transfer.

Especially children with weaker reading skills benefited from videos. This is important, as recent PISA results indicate a simultaneous decline in children's science and reading performance (OECD, 2023, p. 158). Moreover, videos reduced variability in studying rates and strategies, speeding up slower readers and scaffolding learners to cover the whole content. Therefore, videos allow the teacher to influence how study time is distributed across the content and standardize the pace of learning, which helps with both learning and classroom management. These results make videos an equitable learning solution, which benefits the weakest students while not harming skilled students.

Overall, the results confirm that videos can improve science learning for most fifth and sixth graders, and can therefore be recommended for primary education. However, such general guidelines on multimedia instruction are problematic because of the multiple known and possible unknown moderators for the modality effect. Therefore, the recommendation only applies to situations similar to the one detailed in this dissertation, which are nevertheless common occurrences in today's classrooms.

There are multiple factors that increase the practical value of this research. First, the environment, materials, and tests reflect the classroom context, supporting generalization to everyday teaching. Second, many digital textbook platforms offer both narrated and video versions alongside illustrated texts, making the choice of modality an important consideration for teachers. Similarly, deciding whether to allow material access during exercises is a daily instructional decision that teachers can make without requiring extra resources. Third, the increasing use of tablets in classrooms supports more individualized video learning, where children have access to similar materials and learner controls to those in this dissertation. Fourth, the time-limited nature of the study matches typical classroom constraints, meaning that learning sessions like the ones in this dissertation can be included within a lesson. Lastly, decoding tests are straightforward to administer and commonly used, enabling teachers to identify students who could benefit the most from video-based learning.

To conclude, a plausible reason for the learning benefits of videos is that the automatic progression of videos was fast compared to the children's relatively slow reading. This faster studying led to more time for reviewing the videos, which likely supported learning. Faster and less effortful information processing through videos likely stems from multiple sources. First, videos avoided split attention, facilitating the studying of words and pictures in parallel. Second, scaffolding provided by automatically progressing videos covered all content systematically and reduced the active effort required to progress in the presentation, which may have helped in both learning and cognitive load.

Theoretical explanations for the modality effect, such as those offered by cognitive load theory and the cognitive theory of multimedia learning, are grounded in the idea of preventing cognitive overload. However, these explanations have been developed within research settings that differ from the primary school classroom in numerous ways, such as: (a) videos and illustrated texts used in the classroom use different learner controls, (b) time for learning is limited regardless of learner controls, (c) children have lower reading skills than university students, (d) children study in their own classroom and group, using their own tablet, and (e) practicing is not disrupted by artificial dual tasks or questionnaires. As prior research has highlighted the sensitivity of the modality effect to contextual changes, the above differences are significant. The findings of this dissertation indicate that avoiding cognitive overload may not be the main explanation for the modality effect in typical primary school classroom contexts. As a result, the theoretical basis for the modality effect in such contexts may need to be reevaluated.

References

Alemdag, E., & Cagiltay, K. (2018). A systematic review of eye tracking research on multimedia learning. *Computers & Education*, *125*, 413–428. https://doi.org/10.1016/j.compedu.2018.06.023

Amadieu, F., Tricot, A., & Mariné, C. (2009). Prior knowledge in learning from a non-linear electronic document: Disorientation and coherence of the reading sequences. *Computers in Human Behavior*, *25*(2), 381–388. https://doi.org/10.1016/j.chb.2008.12.017

Anmarkrud, Ø., Andresen, A., & Bråten, I. (2019). Cognitive load and working memory in multimedia learning: Conceptual and measurement issues. *Educational Psychologist*, *54*(2), 61–83. https://doi.org/10.1080/00461520.2018.1554484

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In *Psychology of Learning and Motivation* (Vol. 2, pp. 89–195). Elsevier. https://doi.org/10.1016/S0079-7421(08)60422-3

Ayres, P. (2006). Using subjective measures to detect variations of intrinsic cognitive load within problems. *Learning and Instruction*, *16*(5), 389–400. https://doi.org/10.1016/j.learninstruc.2006.09.001

Ayres, P., & Sweller, J. (2014). The split-attention principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning* (2nd ed., pp. 206–226). Cambridge University Press. https://doi.org/10.1017/CBO9781139547369.011

Baddeley, A. (1992). Working memory. *Science*, *255*(5044), 556–559. https://doi.org/10.1126/science.1736359

Baddeley, A. (2012). Working memory: Theories, models, and controversies. *Annual Review of Psychology*, *63*(1), 1–29. https://doi.org/10.1146/annurev-psych-120710-100422

Bahrick, H. P. (1984). Semantic memory content in permastore: Fifty years of memory for Spanish learned in school. *Journal of Experimental Psychology: General*, *113*(1), 1–29. https://doi.org/10.1037/0096-3445.113.1.1

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1). https://doi.org/10.18637/jss.v067.i01

Beege, M., Krieglstein, F., & Arnold, C. (2022). How instructors influence learning with instructional videos—The importance of professional appearance and communication. *Computers & Education*, *185*, 104531. https://doi.org/10.1016/j.compedu.2022.104531

Biard, N., Cojean, S., & Jamet, E. (2018). Effects of segmentation and pacing on procedural learning by video. *Computers in Human Behavior*, *89*, 411–417. https://doi.org/10.1016/j.chb.2017.12.002

Biedert, R., Hees, J., Dengel, A., & Buscher, G. (2012). A robust realtime reading-skimming classifier. *Proceedings of the Symposium on Eye Tracking Research and Applications*, 123–130. https://doi.org/10.1145/2168556.2168575

Brady, T. F., Konkle, T., Alvarez, G. A., & Oliva, A. (2008). Visual long-term memory has a massive storage capacity for object details. *Proceedings of the National Academy of Sciences*, *105*(38), 14325–14329. https://doi.org/10.1073/pnas.0803390105

Breznitz, Z., & Berman, L. (2003). The underlying factors of word reading rate. *Educational Psychology Review*, *15*(3), 247–265. https://doi.org/10.1023/A:1024696101081

Brünken, R., Plass, J. L., & Leutner, D. (2004). Assessment of cognitive load in multimedia learning with dual-task methodology: Auditory load and modality effects. *Instructional Science*, *32*(1/2), 115–132. https://doi.org/10.1023/B:TRUC.0000021812.96911.c5

Brünken, R., Steinbacher, S., Plass, J. L., & Leutner, D. (2002). Assessment of cognitive load in multimedia learning using dual-task methodology. *Experimental Psychology*, *49*(2), 109–119. https://doi.org/10.1027//1618-3169.49.2.109

Brysbaert, M. (2019). How many words do we read per minute? A review and meta-analysis of reading rate. *Journal of Memory and Language*, *109*, 104047. https://doi.org/10.1016/j.jml.2019.104047

Butcher, K. R. (2014). The multimedia principle. In R. E. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning* (2nd ed., pp. 174–205). Cambridge University Press. https://doi.org/10.1017/CBO9781139547369.010

Chambers, C. T., & Johnston, C. (2002). Developmental differences in children's use of rating scales. *Journal of Pediatric Psychology*, *27*(1), 27–36. https://doi.org/10.1093/jpepsy/27.1.27

Chandler, P., & Sweller, J. (1992). The split-attention effect as a factor in the design of instruction. *British Journal of Educational Psychology*, *62*(2), 233–246. https://doi.org/10.1111/j.2044-8279.1992.tb01017.x

Chang, Y., & Choi, S. (2014). Effects of seductive details evidenced by gaze duration. *Neurobiology of Learning and Memory*, *109*, 131–138. https://doi.org/10.1016/j.nlm.2014.01.005

Chen, F., Zhou, J., Wang, Y., Yu, K., Arshad, S. Z., Khawaji, A., & Conway, D. (2016). *Robust multimodal cognitive load measurement*. Springer International Publishing. https://doi.org/10.1007/978-3-319-31700-7

Chen, S. Y., Fan, J.-P., & Macredie, R. D. (2006). Navigation in hypermedia learning systems: Experts vs. novices. *Computers in Human Behavior*, *22*(2), 251–266. https://doi.org/10.1016/j.chb.2004.06.004

Cheon, J., Crooks, S., & Chung, S. (2014). Does segmenting principle counteract modality principle in instructional animation? *British Journal of Educational Technology*, *45*(1), 56–64. https://doi.org/10.1111/bjet.12021

Choi, H.-H., van Merriënboer, J. J. G., & Paas, F. (2014). Effects of the physical environment on cognitive load and learning: Towards a new model of cognitive load. *Educational Psychology Review*, *26*(2), 225–244. https://doi.org/10.1007/s10648-014-9262-6

Clinton-Lisell, V. (2022). Listening ears or reading eyes: A meta-analysis of reading and listening comprehension comparisons. *Review of Educational Research*, *92*(4), 543–582. https://doi.org/10.3102/00346543211060871

Corbalan, G., Kester, L., & van Merriënboer, J. J. G. (2009). Combining shared control with variability over surface features: Effects on transfer test performance and task involvement. *Computers in Human Behavior*, *25*(2), 290–298. https://doi.org/10.1016/j.chb.2008.12.009

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, *24*(1), 87–114. https://doi.org/10.1017/S0140525X01003922

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, *11*(6), 671–684. https://doi.org/10.1016/S0022-5371(72)80001-X

Cycowicz, Y. M. (2019). Orienting and memory to unexpected and/or unfamiliar visual events in children and adults. *Developmental Cognitive Neuroscience*, *36*, 100615. https://doi.org/10.1016/j.dcn.2019.100615

Darwin, C. J., Turvey, M. T., & Crowder, R. G. (1972). An auditory analogue of the sperling partial report procedure: Evidence for brief auditory storage. *Cognitive Psychology*, *3*(2), 255–267. https://doi.org/10.1016/0010-0285(72)90007-2

de Boer, J., Kommers, P. A. M., & de Brock, B. (2011). Using learning styles and viewing styles in streaming video. *Computers & Education*, *56*(3), 727–735. https://doi.org/10.1016/j.compedu.2010.10.015

Dirix, N., Vander Beken, H., De Bruyne, E., Brysbaert, M., & Duyck, W. (2020). Reading text when studying in a second language: An eye-tracking study. *Reading Research Quarterly*, *55*(3), 371–397. https://doi.org/10.1002/rrq.277

Domagk, S., Schwartz, R. N., & Plass, J. L. (2010). Interactivity in multimedia learning: An integrated model. *Computers in Human Behavior*, *26*(5), 1024–1033. https://doi.org/10.1016/j.chb.2010.03.003

Ehri, L. C. (2005). Learning to Read Words: Theory, Findings, and Issues. *Scientific Studies of Reading*, *9*(2), 167–188. https://doi.org/10.1207/s1532799xssr0902_4

Evans, G. W. (2006). Child development and the physical environment. *Annual Review of Psychology*, *57*(1), 423–451. https://doi.org/10.1146/annurev.psych.57.102904.190057

Eysenck, M. W., & Calvo, M. G. (1992). Anxiety and performance: The processing efficiency theory. *Cognition & Emotion*, *6*(6), 409–434. https://doi.org/10.1080/02699939208409696

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/BF03193146

Finnish National Board on Research Integrity TENK. (2019). *The ethical principles of research with human participants and ethical review in the human sciences in Finland* (No. 3; Finnish National Board on Research Integrity TENK Guidelines 2019). https://tenk.fi/sites/default/files/2021-01/Ethical_review_in_human_sciences_2020.pdf

Fiorella, L., & Mayer, R. E. (2021). The Generative Activity Principle in Multimedia Learning. In R. E. Mayer & L. Fiorella (Eds.), *The Cambridge Handbook of Multimedia Learning* (3rd ed., pp. 339–350). Cambridge University Press. https://doi.org/10.1017/9781108894333.036

Fox, J. (2020). *Regression diagnostics second edition*. SAGE Publications, Inc. https://doi.org/10.4135/9781071878651

Friard, O., & Gamba, M. (2016). BORIS: A free, versatile open-source event-logging software for video/audio coding and live observations. *Methods in Ecology and Evolution*, *7*(11), 1325–1330. https://doi.org/10.1111/2041-210X.12584

Friso-van den Bos, I., & van de Weijer-Bergsma, E. (2020). Classroom versus individual working memory assessment: Predicting academic achievement and the role of attention and response inhibition. *Memory*, *28*(1), 70–82. https://doi.org/10.1080/09658211.2019.1682170

Fyfield, M. (2022). YouTube in the secondary classroom: How teachers use instructional videos in mainstream classrooms. *Technology, Pedagogy and Education*, *31*(2), 185–197. https://doi.org/10.1080/1475939X.2021.1980429

Gerjets, P., Scheiter, K., & Cierniak, G. (2009). The scientific value of cognitive load theory: A research agenda based on the structuralist view of theories. *Educational Psychology Review*, *21*(1), 43–54. https://doi.org/10.1007/s10648-008-9096-1

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

Goldman, S. R., & Saul, E. U. (1990). Flexibility in text processing: A strategy competition model. *Learning and Individual Differences*, *2*(2), 181–219. https://doi.org/10.1016/1041-6080(90)90022-9

Gopher, D., & Braune, R. (1984). On the psychophysics of workload: Why bother with subjective measures? *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *26*(5), 519–532. https://doi.org/10.1177/001872088402600504

Gough, P. B., & Tunmer, W. E. (1986). Decoding, reading, and reading disability. *Remedial and Special Education*, *7*(1), 6–10. https://doi.org/10.1177/074193258600700104

Green, P., & MacLeod, C. J. (2016). SIMR: An R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution*, *7*(4), 493–498. https://doi.org/10.1111/2041-210X.12504

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

Haavisto, M., & Jaakkola, T. (2025). Fifth graders study videos faster than illustrated texts: Studying rate, review strategies, and learning implications. *PsyArxiv*. https://doi.org/10.31234/osf.io/mztpw_v1

Haavisto, M., Jaakkola, T., & Lepola, J. (2023). Video outperforms illustrated text: Do old explanations for the modality effect apply in a learner-paced fifth-grade classroom context? *Computers & Education*, 199, 104775. https://doi.org/10.1016/j.compedu.2023.104775

Haavisto, M., Lepola, J., & Jaakkola, T. (2025). The “simple” view of learning from illustrated texts and videos. *Learning and Instruction*, 100, 102200. https://doi.org/10.1016/j.learninstruc.2025.102200

Hannus, M., & Hyönä, J. (1999). Utilization of illustrations during learning of science textbook passages among low- and high-ability children. *Contemporary Educational Psychology*, *24*(2), 95–123. https://doi.org/10.1006/ceps.1998.0987

Hare, C., Johnson, B., Vlahiotis, M., Panda, E. J., Tekok-Kilic, A., & Curtin, S. (2024). Children’s reading outcomes in digital and print mediums: A systematic review. *Journal of Research in Reading*, *47*(3), 309–329. https://doi.org/10.1111/1467-9817.12461

Hasler, B. S., Kersten, B., & Sweller, J. (2007). Learner control, cognitive load and instructional animation. *Applied Cognitive Psychology*, *21*(6), 713–729. https://doi.org/10.1002/acp.1345

Hatsidimitris, G., & Kalyuga, S. (2013). Guided self-management of transient information in animations through pacing and sequencing strategies. *Educational Technology Research and Development*, *61*(1), 91–105. https://doi.org/10.1007/s11423-012-9276-z

Hayden, E., Hiebert, E. H., & Trainin, G. (2019). Patterns of Silent Reading Rate and Comprehension as a Function of Developmental Status, Genre, and Text Position. *Reading Psychology*, *40*(8), 731–767. https://doi.org/10.1080/02702711.2019.1673602

Heiman, G. W., Leo, R. J., Leighbody, G., & Bowler, K. (1986). Word intelligibility decrements and the comprehension of time-compressed speech. *Perception & Psychophysics*, *40*(6), 407–411. https://doi.org/10.3758/BF03208200

Herrlinger, S., Höffler, T. N., Opfermann, M., & Leutner, D. (2017). When do pictures help learning from expository text? Multimedia and modality effects in primary schools. *Research in Science Education*, *47*(3), 685–704. https://doi.org/10.1007/s11165-016-9525-y

Hiebert, E. H., & Daniel, M. (2019). Comprehension and rate during silent reading: Why do some students do poorly? *Reading and Writing*, *32*(7), 1795–1818. https://doi.org/10.1007/s11145-018-9917-7

Higgins, S., Hall, E., Wall, K., Woolner, P., & Mccaughey, C. (2005). *The impact of school environments: A literature review*. The Design Council. http://www.design-council.org.uk/resources/assets/assets/pdf/Publications/The%20Impact%

Höffler, T. N., & Schwartz, R. N. (2011). Effects of pacing and cognitive style across dynamic and non-dynamic representations. *Computers & Education*, *57*(2), 1716–1726. https://doi.org/10.1016/j.compedu.2011.03.012

Holopainen, E. (2003). *Kuullun ja luetun tekstin ymmärtämisstrategiat ja -vaikeudet peruskoulun kolmannella ja yhdeksännellä luokalla* (Jyväskylä Studies in Education, Psychology and Social Research, No. 218) [Doctoral dissertation, University of Jyväskylä]. University of Jyväskylä.

Huang, Y.-M., Shadiev, R., & Hwang, W.-Y. (2016). Investigating the effectiveness of speech-to-text recognition applications on learning performance and cognitive load. *Computers & Education*, *101*, 15–28. https://doi.org/10.1016/j.compedu.2016.05.011

Hyönä, J. (1994). Processing of topic shifts by adults and children. *Reading Research Quarterly*, *29*(1), 76. https://doi.org/10.2307/747739

Hyönä, J. (1995). An eye movement analysis of topic-shift effect during repeated reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(5), 1365. https://doi.org/10.1037/0278-7393.21.5.A

Hyönä, J., Lorch, R. F., & Kaakinen, J. K. (2002). Individual differences in reading to summarize expository text: Evidence from eye fixation patterns. *Journal of Educational Psychology*, *94*(1), 44–55. https://doi.org/10.1037/0022-0663.94.1.44

Hyönä, J., & Niemi, P. (1990). Eye movements during repeated reading of a text. *Acta Psychologica*, *73*(3), 259–280. https://doi.org/10.1016/0001-6918(90)90026-C

Hyvönen, T., Niemi-Sampan, P., & Koli, A. (2018). *Materiaalit ja tekijänoikeudet opetuksessa*. kopiosto.fi/app/uploads/2018/08/01112400/Materiaalit_ja_tekijanoikeudet_opetuksessa_selvitys_2018.pdf

Inan, F. A., Crooks, S. M., Cheon, J., Ari, F., Flores, R., Kurucay, M., & Paniukov, D. (2015). The reverse modality effect: Examining student learning from interactive computer-based instruction: Multimedia design principles. *British Journal of Educational Technology*, *46*(1), 123–130. https://doi.org/10.1111/bjet.12129

Jiang, D., & Kalyuga, S. (2020). Confirmatory factor analysis of cognitive load ratings supports a two-factor model. *The Quantitative Methods for Psychology*, *16*(3), 216–225. https://doi.org/10.20982/tqmp.16.3.p216

Kalyuga, S. (2007). Expertise reversal effect and its implications for learner-tailored instruction. *Educational Psychology Review*, *19*(4), 509–539. https://doi.org/10.1007/s10648-007-9054-3

Kalyuga, S. (2011). Cognitive load theory: How many types of load does it really need? *Educational Psychology Review*, *23*(1), 1–19. https://doi.org/10.1007/s10648-010-9150-7

Kalyuga, S. (2014). The expertise reversal principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning* (2nd ed., pp. 576–597). Cambridge University Press. https://doi.org/10.1017/CBO9781139547369.028

Kalyuga, S., Chandler, P., & Sweller, J. (1998). Levels of Expertise and Instructional Design. *Human Factors*, *40*(1), 1–17. https://doi.org/10.1518/001872098779480587

Kay, R. H. (2012). Exploring the use of video podcasts in education: A comprehensive review of the literature. *Computers in Human Behavior*, *28*(3), 820–831. https://doi.org/10.1016/j.chb.2012.01.011

Kim, J., Guo, P. J., Seaton, D. T., Mitros, P., Gajos, K. Z., & Miller, R. C. (2014). Understanding in-video dropouts and interaction peaks in online lecture videos. *Proceedings of the First ACM Conference on Learning @ Scale Conference*, 31–40. https://doi.org/10.1145/2556325.2566237

Kim, Y.-S., & Phillips, B. (2014). Cognitive Correlates of Listening Comprehension. *Reading Research Quarterly*, *49*(3), 269–281. https://doi.org/10.1002/rrq.74

Kingstone, A., Smilek, D., & Eastwood, J. D. (2008). Cognitive ethology: A new approach for studying human cognition. *British Journal of Psychology*, *99*(3), 317–340. https://doi.org/10.1348/000712607X251243

Kintsch, W., & Van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review*, *85*(5), 363–394. https://doi.org/10.1037/0033-295X.85.5.363

Kirk, R. E. (2009). Experimental design. In R. Millsap & A. Maydeu-Olivares, *The SAGE Handbook of Quantitative Methods in Psychology* (pp. 23–45). SAGE Publications Ltd. https://doi.org/10.4135/9780857020994.n2

Klatte, M., Lachmann, T., & Meis, M. (2010). Effects of noise and reverberation on speech perception and listening comprehension of children and adults in a classroom-like setting. *Noise and Health*, *12*(49), 270. https://doi.org/10.4103/1463-1741.70506

Klein, J. (2022). Improving the reproducibility of findings by updating research methodology. *Quality & Quantity*, *56*(3), 1597–1609. https://doi.org/10.1007/s11135-021-01196-6

Knoop-van Campen, C. A. N., Segers, E., & Verhoeven, L. (2018). The modality and redundancy effects in multimedia learning in children with dyslexia. *Dyslexia*, *24*(2), 140–155. https://doi.org/10.1002/dys.1585

Knoop-van Campen, C. A. N., Segers, E., & Verhoeven, L. (2019). Modality and redundancy effects, and their relation to executive functioning in children with dyslexia. *Research in Developmental Disabilities*, *90*, 41–50. https://doi.org/10.1016/j.ridd.2019.04.007

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, *15*(2), 155–163. https://doi.org/10.1016/j.jcm.2016.02.012

Kuperman, V., Kyröläinen, A.-J., Porretta, V., Brysbaert, M., & Yang, S. (2021). A lingering question addressed: Reading rate and most efficient listening rate are highly similar. *Journal of*

Experimental Psychology: Human Perception and Performance, *47*(8), 1103–1112. https://doi.org/10.1037/xhp0000932

Lakens, D., Scheel, A. M., & Isager, P. M. (2018). Equivalence testing for psychological research: A tutorial. *Advances in Methods and Practices in Psychological Science*, *1*(2), 259–269. https://doi.org/10.1177/2515245918770963

Language and Reading Research Consortium. (2015). Learning to read: Should we keep things simple? *Reading Research Quarterly*, *50*(2), 151–169. https://doi.org/10.1002/rrq.99

Le, Y., Chen, Z., Liu, S., Pang, W., & Deng, C. (2021). Investigating the effectiveness of emotional design principle to attenuate ego depletion effect. *Computers & Education*, *174*, 104311. https://doi.org/10.1016/j.compedu.2021.104311

Leahy, W., Chandler, P., & Sweller, J. (2003). When auditory presentations should and should not be a component of multimedia instruction. *Applied Cognitive Psychology*, *17*(4), 401–418. https://doi.org/10.1002/acp.877

Leahy, W., & Sweller, J. (2011). Cognitive load theory, modality of presentation and the transient information effect. *Applied Cognitive Psychology*, *25*(6), 943–951. https://doi.org/10.1002/acp.1787

Leahy, W., & Sweller, J. (2016). Cognitive load theory and the effects of transient information on the modality effect. *Instructional Science*, *44*(1), 107–123. https://doi.org/10.1007/s11251-015-9362-9

Lee, H. Y., & List, A. (2019). Processing of texts and videos: A strategy-focused analysis. *Journal of Computer Assisted Learning*, *35*(2), 268–282. https://doi.org/10.1111/jcal.12328

Leppink, J., Paas, F., Van der Vleuten, C. P. M., Van Gog, T., & Van Merriënboer, J. J. G. (2013). Development of an instrument for measuring different types of cognitive load. *Behavior Research Methods*, *45*(4), 1058–1072. https://doi.org/10.3758/s13428-013-0334-1

Li, L., Wang, X., & Wallace, M. P. (2022). I determine my learning path, or not? A study of different learner control conditions in online video-based learning. *Frontiers in Psychology*, *13*. https://www.frontiersin.org/articles/10.3389/fpsyg.2022.973758

Lindeman, J. (1998). *ALLU. Ala-asteen lukutesti. Tekniset tiedot.* Turun yliopisto, Oppimistutkimuksen keskus.

List, A., & Ballenger, E. E. (2019). Comprehension across mediums: The case of text and video. *Journal of Computing in Higher Education*, *31*(3), 514–535. https://doi.org/10.1007/s12528-018-09204-9

Liversedge, S. P., Drieghe, D., Li, X., Yan, G., Bai, X., & Hyönä, J. (2016). Universality in eye movements and reading: A trilingual investigation. *Cognition*, *147*, 1–20. https://doi.org/10.1016/j.cognition.2015.10.013

Low, R., & Sweller, J. (2014). The modality principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning* (2nd ed., pp. 227–246). Cambridge University Press. https://doi.org/10.1017/CBO9781139547369.012

Lüdecke, D., Ben-Shachar, M., Patil, I., Waggoner, P., & Makowski, D. (2021). performance: An R package for assessment, comparison and testing of statistical models. *Journal of Open Source Software*, *6*(60), 3139. https://doi.org/10.21105/joss.03139

Marcus, N., Cooper, M., & Sweller, J. (1996). Understanding instructions. *Journal of Educational Psychology*, *88*(1), 49. https://doi.org/10.1037/0022-0663.88.1.49

Marsh, H. W. (1986). Negative item bias in ratings scales for preadolescent children: A cognitive-developmental phenomenon. *Developmental Psychology*, *22*(1), 37–49. https://doi.org/10.1037/0012-1649.22.1.37

Mason, L., Tornatora, M. C., & Pluchino, P. (2013). Do fourth graders integrate text and picture in processing and learning from an illustrated science text? Evidence from eye-movement patterns. *Computers & Education*, *60*(1), 95–109. https://doi.org/10.1016/j.compedu.2012.07.011

Mayer, R. E. (2014). Cognitive theory of multimedia learning. In R. E. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning* (2nd ed., pp. 43–71). Cambridge University Press. https://doi.org/10.1017/CBO9781139547369.005

Mayer, R. E. (2020). *Multimedia learning* (3rd ed.). Cambridge University Press. https://doi.org/10.1017/9781316941355

Mayer, R. E., & Fiorella, L. (2014). Principles for reducing extraneous processing in multimedia learning: Coherence, signaling, redundancy, spatial contiguity, and temporal contiguity principles. In R. E. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning* (2nd ed., pp. 279–315). Cambridge University Press. https://doi.org/10.1017/CBO9781139547369.015

Mayer, R. E., & Gallini, J. K. (1990). When is an illustration worth ten thousand words? *Journal of Educational Psychology*, *82*(4), 715–726. https://doi.org/10.1037/0022-0663.82.4.715

Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist*, *38*(1), 43–52. https://doi.org/10.1207/S15326985EP3801_6

Merkt, M., Ballmann, A., Felfeli, J., & Schwan, S. (2018). Pauses in educational videos: Testing the transience explanation against the structuring explanation. *Computers in Human Behavior*, *89*, 399–410. https://doi.org/10.1016/j.chb.2018.01.013

Merkt, M., Hoppe, A., Bruns, G., Ewerth, R., & Huff, M. (2022). Pushing the button: Why do learners pause online videos? *Computers & Education*, *176*, 104355. https://doi.org/10.1016/j.compedu.2021.104355

Merkt, M., Weigand, S., Heier, A., & Schwan, S. (2011). Learning with videos vs. learning with print: The role of interactive features. *Learning and Instruction*, S0959475211000247. https://doi.org/10.1016/j.learninstruc.2011.03.004

Michel, E., Roebers, C. M., & Schneider, W. (2007). Educational films in the classroom: Increasing the benefit. *Learning and Instruction*, *17*(2), 172–183. https://doi.org/10.1016/j.learninstruc.2007.01.005

Moreno, R., & Mayer, R. E. (1999). Cognitive principles of multimedia learning: The role of modality and contiguity. *Journal of Educational Psychology*, *91*(2), 358–368. https://doi.org/10.1037/0022-0663.91.2.358

Mutlu-Bayraktar, D., Cosgun, V., & Altan, T. (2019). Cognitive load in multimedia learning environments: A systematic review. *Computers & Education*, *141*, 103618. https://doi.org/10.1016/j.compedu.2019.103618

Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R^2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, *4*(2), 133–142. https://doi.org/10.1111/j.2041-210x.2012.00261.x

Nevala, J., & Lyytinen, H. (2000). *Sanaketjutesti*. Niilo Mäki Instituutti & Jyväskylän yliopiston lapsitutkimuskeskus.

O'Brien, R. M. (2007). A caution regarding rules of thumb for variance inflation factors. *Quality & Quantity*, *41*(5), 673–690. https://doi.org/10.1007/s11135-006-9018-6

OECD. (2023). *PISA 2022 Results (Volume I): The State of Learning and Equity in Education*. OECD. https://doi.org/10.1787/53f23881-en

Official Statistics of Finland. (2021). *Students and qualifications*. https://www.stat.fi/til/opiskt/meta_en.html

Paas, F. (1992). Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive-load approach. *Journal of Educational Psychology*, *84*(4), 429–434. https://doi.org/10.1037/0022-0663.84.4.429

Paas, F., & Sweller, J. (2012). An evolutionary upgrade of cognitive load theory: Using the human motor system and collaboration to support the learning of complex cognitive tasks. *Educational Psychology Review*, *24*(1), 27–45. https://doi.org/10.1007/s10648-011-9179-2

Paas, F., & Sweller, J. (2014). Implications of cognitive load theory for multimedia learning. In R. E. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning* (2nd ed., pp. 27–42). Cambridge University Press. https://doi.org/10.1017/CBO9781139547369.004

Paas, F., Tuovinen, J. E., Tabbers, H., & Van Gerven, P. W. M. (2003). Cognitive load measurement as a means to advance cognitive load theory. *Educational Psychologist*, *38*(1), 63–71. https://doi.org/10.1207/S15326985EP3801_8

Paas, F., & Van Merriënboer, J. J. G. (1993). The efficiency of instructional conditions: An approach to combine mental effort and performance measures. *Human Factors*, *35*(4), 737–743. https://doi.org/10.1177/001872089303500412

Paas, F., van Merriënboer, J. J. G., & Adam, J. J. (1994). Measurement of cognitive load in instructional research. *Perceptual and Motor Skills*, *79*(1), 419–430. https://doi.org/10.2466/pms.1994.79.1.419

Paivio, A. (1990). *Mental Representations: A dual coding approach*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195066661.001.0001

Park, B., Flowerday, T., & Brünken, R. (2015). Cognitive and affective effects of seductive details in multimedia learning. *Computers in Human Behavior*, *44*, 267–278. https://doi.org/10.1016/j.chb.2014.10.061

Park, B., Moreno, R., Seufert, T., & Brünken, R. (2011). Does cognitive load moderate the seductive details effect? A multimedia study. *Computers in Human Behavior*, *27*(1), 5–10. https://doi.org/10.1016/j.chb.2010.05.006

Pastore, R. (2010). The effects of diagrams and time-compressed instruction on learning and learners' perceptions of cognitive load. *Educational Technology Research and Development*, *58*(5), 485–505. https://doi.org/10.1007/s11423-009-9145-6

Pastore, R. (2012). The effects of time-compressed instruction and redundancy on learning and learners' perceptions of cognitive load. *Computers & Education*, *58*(1), 641–651. https://doi.org/10.1016/j.compedu.2011.09.018

Pattier, D. (2021). Teachers and YouTube: The use of video as an educational resource. *Ricerche Di Pedagogia e Didattica. Journal of Theories and Research in Education*, *16*(1), Article 1. https://doi.org/10.6092/issn.1970-2221/11584

Penney, C. G. (1989). Modality effects and the structure of short-term verbal memory. *Memory & Cognition*, *17*(4), 398–422. https://doi.org/10.3758/BF03202613

Perfetti, C. (1985). *Reading ability*. Oxford University Press.

Perfetti, C., & Stafura, J. (2014). Word Knowledge in a Theory of Reading Comprehension. *Scientific Studies of Reading*, *18*(1), 22–37. https://doi.org/10.1080/10888438.2013.827687

Peterson, L., & Peterson, M. J. (1959). Short-term retention of individual verbal items. *Journal of Experimental Psychology*, *58*(3), 193–198. https://doi.org/10.1037/h0049234

Pustejovsky, J. E. (2022). *clubSandwich: Cluster-robust (sandwich) variance estimators with small-sample corrections* [Computer software]. https://cran.r-project.org/web/packages/clubSandwich/

Pustejovsky, J. E., & Tipton, E. (2018). Small-sample methods for cluster-robust variance estimation and hypothesis testing in fixed effects models. *Journal of Business & Economic Statistics*, *36*(4), 672–683. https://doi.org/10.1080/07350015.2016.1247004

R Core Team. (2024). *R: A language and environment for statistical computing* [Computer software]. R Foundation for Statistical Computing. https://www.r-project.org/

Rasch, T., & Schnotz, W. (2009). Interactive and non-interactive pictures in multimedia learning environments: Effects on learning outcomes and learning efficiency. *Learning and Instruction*, *19*(5), 411–422. https://doi.org/10.1016/j.learninstruc.2009.02.008

Reinwein, J. (2012). Does the modality effect exist? And if so, which modality effect? *Journal of Psycholinguistic Research*, *41*(1), 1–32. https://doi.org/10.1007/s10936-011-9180-4

Reinwein, J., & Tassé, S. (2022). Modality effects examined by means of an online sentence-picture comparison task. *Journal of Psycholinguistic Research*, *51*(3), 521–542. https://doi.org/10.1007/s10936-022-09849-9

Rey, G. D. (2012). A review of research and a meta-analysis of the seductive detail effect. *Educational Research Review*, *7*(3), 216–237. https://doi.org/10.1016/j.edurev.2012.05.003

Rey, G. D., Beege, M., Nebel, S., Wirzberger, M., Schmitt, T. H., & Schneider, S. (2019). A meta-analysis of the segmenting effect. *Educational Psychology Review*, *31*(2), 389–419. https://doi.org/10.1007/s10648-018-9456-4

Ritzhaupt, A. D., Gomes, N. D., & Barron, A. E. (2008). The effects of time-compressed audio and verbal redundancy on learner performance and satisfaction. *Computers in Human Behavior*, *24*(5), 2434–2445. https://doi.org/10.1016/j.chb.2008.02.017

Roy-Charland, A., Saint-Aubin, J., & Evans, M. A. (2007). Eye movements in shared book reading with children from kindergarten to Grade 4. *Reading and Writing*, *20*(9), 909–931. https://doi.org/10.1007/s11145-007-9059-9

Rummer, R., Schweppe, J., Fürstenberg, A., Scheiter, K., & Zindler, A. (2011). The perceptual basis of the modality effect in multimedia learning. *Journal of Experimental Psychology: Applied*, *17*(2), 159–173. https://doi.org/10.1037/a0023588

Scheiter, K., & Gerjets, P. (2007). Learner control in hypermedia environments. *Educational Psychology Review*, *19*(3), 285–307. https://doi.org/10.1007/s10648-007-9046-3

Scheiter, K., Schüler, A., Gerjets, P., Huk, T., & Hesse, F. W. (2014). Extending multimedia research: How do prerequisite knowledge and reading comprehension affect learning from text and pictures. *Computers in Human Behavior*, *31*, 73–84. https://doi.org/10.1016/j.chb.2013.09.022

Schmeck, A., Opfermann, M., van Gog, T., Paas, F., & Leutner, D. (2015). Measuring cognitive load with subjective rating scales during problem solving: Differences between immediate and delayed ratings. *Instructional Science*, *43*(1), 93–114. https://doi.org/10.1007/s11251-014-9328-3

Schmuckler, M. A. (2001). What is ecological validity? A dimensional analysis. *Infancy*, *2*(4), 419–436. https://doi.org/10.1207/S15327078IN0204_02

Schnotz, W. (2014). Integrated model of text and picture comprehension. In R. E. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning* (2nd ed., pp. 72–103). Cambridge University Press. https://doi.org/10.1017/CBO9781139547369.006

Schüler, A., Scheiter, K., Rummer, R., & Gerjets, P. (2012). Explaining the modality effect in multimedia learning: Is it due to a lack of temporal contiguity with written text and pictures? *Learning and Instruction*, *22*(2), 92–102. https://doi.org/10.1016/j.learninstruc.2011.08.001

Schüler, A., Scheiter, K., & van Genuchten, E. (2011). The role of working memory in multimedia instruction: Is working memory working during learning from text and pictures? *Educational Psychology Review*, *23*(3), 389–411. https://doi.org/10.1007/s10648-011-9168-5

Segers, E., Verhoeven, L., & Hulstijn-Hendrikse, N. (2008). Cognitive processes in children's multimedia text learning. *Applied Cognitive Psychology*, *22*(3), 375–387. https://doi.org/10.1002/acp.1413

Simola, J., Salojärvi, J., & Kojo, I. (2008). Using hidden Markov model to uncover processing states from eye movements in information search tasks. *Cognitive Systems Research*, *9*(4), 237–251. https://doi.org/10.1016/j.cogsys.2008.01.002

Singh, A.-M., Marcus, N., & Ayres, P. (2012). The transient information effect: Investigating the impact of segmentation on spoken and written text. *Applied Cognitive Psychology*, *26*(6), 848–853. https://doi.org/10.1002/acp.2885

Smith, R., Snow, P., Serry, T., & Hammond, L. (2021). The Role of Background Knowledge in Reading Comprehension: A Critical Review. *Reading Psychology*, *42*(3), 214–240. https://doi.org/10.1080/02702711.2021.1888348

Smith, S. M., & Vela, E. (2001). Environmental context-dependent memory: A review and meta-analysis. *Psychonomic Bulletin & Review*, *8*(2), 203–220. https://doi.org/10.3758/BF03196157

Snell, J., & Grainger, J. (2019). Readers are parallel processors. *Trends in Cognitive Sciences*, *23*(7), 537–546. https://doi.org/10.1016/j.tics.2019.04.006

Spanjers, I. A. E., van Gog, T., & van Merriënboer, J. J. G. (2010). A theoretical analysis of how segmentation of dynamic visualizations optimizes students' learning. *Educational Psychology Review*, *22*(4), 411–423. https://doi.org/10.1007/s10648-010-9135-6

Spanjers, I. A. E., Van Gog, T., Wouters, P., & Van Merriënboer, J. J. G. (2012). Explaining the segmentation effect in learning from animations: The role of pausing and temporal cueing. *Computers & Education*, *59*(2), 274–280. https://doi.org/10.1016/j.compedu.2011.12.024

Spanjers, I. A. E., Wouters, P., van Gog, T., & van Merriënboer, J. J. G. (2011). An expertise reversal effect of segmentation in learning from animated worked-out examples. *Computers in Human Behavior*, *27*(1), 46–52. https://doi.org/10.1016/j.chb.2010.05.011

Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs: General and Applied*, *74*(11), 1–29. https://doi.org/10.1037/h0093759

Spichtig, A. N., Hiebert, E. H., Vorstius, C., Pascoe, J. P., David Pearson, P., & Radach, R. (2016). The decline of comprehension-based silent reading efficiency in the United States: A comparison of current data with performance in 1960. *Reading Research Quarterly*, *51*(2), 239–259. https://doi.org/10.1002/rrq.137

Spichtig, A. N., Pascoe, J. P., Gehsmann, K. M., Gu, F., & Ferrara, J. D. (2022). The Interaction of Silent Reading Rate, Academic Vocabulary, and Comprehension Among Students in Grades 2–12. *Reading Research Quarterly*, *57*(3), 1003–1019. https://doi.org/10.1002/rrq.457

Sweller, J. (2010). Element interactivity and intrinsic, extraneous, and germane cognitive load. *Educational Psychology Review*, *22*(2), 123–138. https://doi.org/10.1007/s10648-010-9128-5

Sweller, J. (2020). Cognitive load theory and educational technology. *Educational Technology Research and Development*, *68*(1), 1–16. https://doi.org/10.1007/s11423-019-09701-3

Sweller, J., van Merriënboer, J. J. G., & Paas, F. (2019). Cognitive architecture and instructional design: 20 years later. *Educational Psychology Review*, *31*(2), 261–292. https://doi.org/10.1007/s10648-019-09465-5

Tabbers, H. K., & de Koeijer, B. (2010). Learner control in animated multimedia instructions. *Instructional Science*, *38*(5), 441–453. https://doi.org/10.1007/s11251-009-9119-4

Tabbers, H. K., Martens, R. L., & Van Merrienboer, J. J. G. (2001). The modality effect in multimedia instructions. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *23*(23). https://escholarship.org/uc/item/8fd0d9x5

Tabbers, H. K., Martens, R. L., & Van Merrienboer, J. J. G. (2004). Multimedia instructions and cognitive load theory: Effects of modality and cueing. *British Journal of Educational Psychology*, *74*(1), 71–81. https://doi.org/10.1348/000709904322848824

Taber, K. S. (2018). The use of Cronbach's alpha when developing and reporting research instruments in science education. *Research in Science Education*, *48*(6), 1273–1296. https://doi.org/10.1007/s11165-016-9602-2

Tarchi, C., Zaccoletti, S., & Mason, L. (2021). Learning from text, video, or subtitles: A comparative analysis. *Computers & Education*, *160*, 104034. https://doi.org/10.1016/j.compedu.2020.104034

Taylor, S. E. (1965). Eye Movements in Reading: Facts and Fallacies. *American Educational Research Journal*, *2*(4), 187–202. https://doi.org/10.3102/00028312002004187

ten Hove, P., & van der Meij, H. (2015). Like it or not. What characterizes YouTube's more popular instructional videos? *Technical Communication*, *62*, 48–62.

Torppa, Georgiou, Lerkkanen, Niemi, Poikkeus, & Nurmi. (2016). Examining the simple view of reading in a transparent orthography: A longitudinal study from kindergarten to grade 3. *Merrill-Palmer Quarterly*, *62*(2), 179. https://doi.org/10.13110/merrpalmquar1982.62.2.0179

Tortorelli, L. S. (2019). Reading rate in informational text: Norms and implications for theory and practice in the primary grades. *Reading Psychology*, *40*(3), 293–324. https://doi.org/10.1080/02702711.2019.1621011

van Gog, T., Kirschner, F., Kester, L., & Paas, F. (2012). Timing and frequency of mental effort measurement: Evidence in favour of repeated measures. *Applied Cognitive Psychology*, *26*(6), 833–839. https://doi.org/10.1002/acp.2883

van Gog, T., & Paas, F. (2008). Instructional efficiency: Revisiting the original construct in educational research. *Educational Psychologist*, *43*(1), 16–26. https://doi.org/10.1080/00461520701756248

Vettori, G., Di Leonardo, L., Secchi, S., Astolfi, A., & Bigozzi, L. (2022). Primary school children's verbal working memory performances in classrooms with different acoustic conditions. *Cognitive Development*, 64, 101256. https://doi.org/10.1016/j.cogdev.2022.101256

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, *48*(4), 817–838. https://doi.org/10.2307/1912934

Witteman, M. J., & Segers, E. (2010). The modality effect tested in children in a user-paced multimedia environment. *Journal of Computer Assisted Learning*, *26*(2), 132–142. https://doi.org/10.1111/j.1365-2729.2009.00335.x

Wolf, M. C., Muijselaar, M. M. L., Boonstra, A. M., & de Bree, E. H. (2019). The relationship between reading and listening comprehension: Shared and modality-specific components. *Reading and Writing*, *32*(7), 1747–1767. https://doi.org/10.1007/s11145-018-9924-8

Wong, A., Leahy, W., Marcus, N., & Sweller, J. (2012). Cognitive load theory, the transient information effect and e-learning. *Learning and Instruction*, *22*(6), 449–457. https://doi.org/10.1016/j.learninstruc.2012.05.004

Woodcock, R. W., & Clark, C. R. (1968). *Comprehension of a narrative passage by elementary school children as a function of listening rate, retention period, and IQ* (No. Volume V, No. 2; IMRID Papers and Reports). https://files.eric.ed.gov/fulltext/ED105413.pdf

Xie, B., & Salvendy, G. (2000). Review and reappraisal of modelling and predicting mental workload in single- and multi-task environments. *Work & Stress*, *14*(1), 74–99. https://doi.org/10.1080/026783700417249

Yang, F.-Y., Chang, C.-Y., Chien, W.-R., Chien, Y.-T., & Tseng, Y.-H. (2013). Tracking learners' visual attention during a multimedia presentation in a real classroom. *Computers & Education*, *62*, 208–220. https://doi.org/10.1016/j.compedu.2012.10.009

Zeileis, A. (2004). Econometric computing with HC and HAC covariance matrix estimators. *Journal of Statistical Software*, *11*, 1–17. https://doi.org/10.18637/jss.v011.i10

Zhu, J., Dawson, K., & Ritzhaupt, A. D. (2020). Investigating how multimedia and modality design principles influence student learning performance, satisfaction, mental effort, and visual attention. *Journal of Educational Multimedia and Hypermedia*, *29*(3), 265–284.

TURUN
YLIOPISTO
UNIVERSITY
OF TURKU

Painosalama, Turku, Finland 2025

ISBN 978-952-02-0470-9 (PRINT)
ISBN 978-952-02-0471-6 (PDF)
ISSN 0082-6987 (Print)
ISSN 2343-3191 (Online)