Computers & Education 199 (2023) 104775

Contents lists available at ScienceDirect

Computers & Education

journal homepage: www.elsevier.com/locate/compedu

Video outperforms illustrated text: Do old explanations for the modality effect apply in a learner-paced fifth-grade classroom context?

Mikko Haavisto [a,*], Tomi Jaakkola [b], Janne Lepola [a]

[a] *Department of Teacher Education, University of Turku, Seminaarinkatu 1, 26100, Rauma, Finland*
[b] *Faculty of Education and Culture, Tampere University, Åkerlundinkatu 5, 33100, Tampere, Finland*

A R T I C L E I N F O

Keywords:
Elementary education
Improving classroom teaching
Media in education
Pedagogical issues
Applications in subject areas

A B S T R A C T

The modality effect occurs when people learn better from a combination of pictures and narration than from a combination of pictures and written text. Despite the strong empirical results in earlier studies, the modality effect has been less prominent in later studies of children in learner-paced settings. However, the generalizability of these results in practice may be limited because the studies included notable differences compared to a classroom context. The present study examined the modality effect in a learner-paced classroom context. In a within-subjects experiment, fifth graders learned from illustrated texts and videos and completed pre-, post-, and delayed tests on two science topics. The video group outperformed the illustrated text group in retention, delayed retention, cognitive load, and efficiency measures but there were no statistical differences in transfer. In both learning conditions, the cognitive load was moderate and did not correlate with any learning outcomes. The results suggest that while the modality effect can occur in a learner-paced classroom context, it may not be based on the avoidance of cognitive overload. Alternative explanations concerning the differences in settings and materials between classroom contexts and modality effect research are discussed.

1. Introduction

The question of how people learn from multimedia instruction has attracted much research interest over the past few decades (see Mayer, 2020, p. xii). One of the most studied areas is the modality effect (Mayer, 2020, p. 291), which occurs when the information presented in multiple sensory modalities (e.g., pictures and narration) enhances learning in comparison with that of a single modality (e.g., pictures and written text; Low & Sweller, 2014). However, studies have shown that the modality effect is sensitive to changes in the learning context, that is, it is influenced (or moderated) by multiple variables (Reinwein, 2012). This has two major implications for research and practice. First, general instructional design suggestions, such as “present pictures with spoken text rather than written text” (i.e., the originally proposed modality principle; Moreno & Mayer, 1999), are unlikely to be helpful to teachers (Scheiter et al., 2014; Schüler et al., 2012). Second, the perspective of ecological validity becomes crucial: if the experimental context of a study differs from an authentic context concerning a known or yet unknown moderator, its results cannot reliably contribute to instructional practice (see Kingstone et al., 2008), which is an important goal of multimedia research (Mayer, 2020, pp. 23–24).

* Corresponding author.
E-mail addresses: mijuhaa@utu.fi (M. Haavisto), tomi.jaakkola@tuni.fi (T. Jaakkola), janlep@utu.fi (J. Lepola).

https://doi.org/10.1016/j.compedu.2023.104775
Received 27 August 2022; Received in revised form 1 February 2023; Accepted 4 March 2023
Available online 6 March 2023
0360-1315/

Illustrated texts (that combine pictures and written text) have arguably been the most common teaching material for centuries, but nowadays, videos (that combine pictures and narration) have become an increasingly popular alternative. This has particularly been the case in primary schools, where digital textbook producers have begun to provide content in video format as an option to illustrated texts, and personal tablet computers enable students to learn from videos at their own pace. For example, an estimated 85% of teachers in grades 1–9 use YouTube videos in their classrooms in Finland (Hyvönen et al., 2018). Despite this, little research has been done on the relative benefits of modern videos and illustrated texts in a classroom context. For instance, only one of the 51 studies in Reinwein's (2012) meta-analysis addressed the modality effect in a learner-paced setting with young learners. Although some more recent studies have been published on this matter (e.g., Knoop-van Campen et al., 2018, 2019; Witteman & Segers, 2010), these results may not generalize to a classroom context because the studies were not conducted in a classroom setting and the narrated materials did not resemble modern videos that provide flexible navigation options and full-screen illustrations. It is therefore unclear whether the substitution of video for illustrated text in primary school classrooms is justified from a learning perspective. This motivated the primary research question of the present study.

RQ1. Does learning from videos and learning from illustrated texts in a classroom context produce different science learning outcomes measured by retention, delayed retention, and transfer?

To address this, the present study investigated the modality effect in a within-subjects experiment, in which fifth graders used videos and illustrated texts to learn two science topics independently at their own pace. The aim was to replicate the use of these materials in an authentic classroom learning situation as closely as possible, without compromising the comparability of the learning conditions.

1.1. Cognitive load theory and the modality effect

The modality effect was derived from cognitive load theory (Leahy & Sweller, 2016), which is the dominant theoretical framework for understanding multimedia learning effects. In this study, cognitive load is measured and discussed from the perspective of the learner (see Kalyuga, 2007, 2011). That is, cognitive load occurs when the learner effortfully processes information. Furthermore, total cognitive load is defined as the sum of intrinsic cognitive load (i.e., cognitive load caused by the to-be-learned information itself) and extraneous cognitive load (i.e., cognitive load caused by information that is unnecessary for the learning process; see Kalyuga, 2011). The extraneous load can be reduced by good instructional design to free up working memory resources for the intrinsic load (given that there is any), which leads to better learning (Paas & Sweller, 2014). Conversely, when the total cognitive load exceeds the learner's cognitive capacity, a cognitive overload occurs, leading to inferior learning (Ginns, 2005).

According to cognitive load theory, learning from an illustrated narration such as a video may cause less cognitive load than learning from an illustrated text, because humans have specifically evolved to understand speech (Paas & Sweller, 2014). In contrast, written text can only be understood through the retrieval of learned letter and word patterns from long-term memory (Paas & Sweller, 2014), which may cause an extraneous load, especially for children, whose basic reading skills are still developing (Reinwein & Tassé, 2022). Another benefit of videos is that the learner can simultaneously look at the pictures while listening to the narration, therefore efficiently receiving both auditory and visual input to verbal and visual working memory (see Baddeley, 1992; Penney, 1989), and preventing an overload in the visual channel (cf. Schüler et al., 2012). In contrast, learning from illustrated texts requires learners to split their attention (Chandler & Sweller, 1992) between the pictures and the text, which may underlie the modality effect (Tabbers et al., 2004; cf., Moreno & Mayer, 1999; Reinwein & Tassé, 2022).

Despite the above benefits, there are also challenges in learning from a video. Namely, information is presented and accessed serially in listening (Brysbaert, 2019), whereas there are indications that multiple words can be processed simultaneously in reading (Snell & Grainger, 2019). Due to its serial nature, spoken text is difficult to present in a way that enables learners to quickly review previous information. Singh et al. (2012) have argued that listening may cause additional cognitive load compared to reading because listeners must hold information in working memory, whereas the readers can simply reread. This leads to the so-called transient information effect in narrations: learning is decreased because the presented information is replaced by new information at a pace that exceeds the learner's cognitive resources (Leahy & Sweller, 2011). While readers can quickly and easily reread and adjust their reading pace in almost any text, providing listeners with effective ways of controlling narrations is the subject of ongoing research (e.g., Hatsidimitris & Kalyuga, 2013; Merkt et al., 2011, 2022), including the present study.

All the above explanations for the modality effect are related to the avoidance of cognitive overload. However, if the intrinsic load is low, cognitive load theory becomes less relevant in explaining the learning effects (Wong et al., 2012), since learners have the extra cognitive capacity to process the potential extraneous load. For example, a child may be able to effortfully recognize words and alternate their attention between pictures and text if sufficient time is available. Despite this, most of the studies of the modality effect in children have not measured cognitive load, even though the avoidance of cognitive overload is given as an explanation for the modality effect (e.g., Knoop-van Campen et al., 2018, 2019; Leahy & Sweller, 2011; Segers et al., 2008; Witteman & Segers, 2010; Wong et al., 2012). Furthermore, studies that do measure cognitive load and learning commonly do not statistically analyze how the former affects the latter (e.g., Huang et al., 2016; Inan et al., 2015; Liu et al., 2021). Thus, the question arises whether the differences in learning in these studies can be attributed to differences in cognitive load, and what in general is the relationship between cognitive load and learning. These considerations motivated the second and third research questions of the present study.

RQ2. Does learning from videos and learning from illustrated texts in a classroom context produce different cognitive loads?

RQ3. What is the relationship between cognitive load and learning outcomes?

The present study sought to answer these questions by measuring cognitive load and by analyzing its relationship to learning outcomes. Moreover, these variables were used to calculate an efficiency metric (Paas & Van Merriënboer, 1993) to show the potential benefits of the learning materials in both increased learning and decreased mental effort. This reflects the fact that cognitive load is also an important outcome measure for teachers to consider: the material that produces better learning outcomes is not always the better option if it makes the learners exhausted.

1.2. Pacing control and learning time as moderators of the modality effect

Most of the prior research on the modality effect has been concerned with studying system-paced conditions, in which learners cannot control the learning materials, rather than learner-paced conditions (Reinwein, 2012). However, system-pacing is not common in individual multimedia learning, as students can read at their own pace and browse previous content in both paper and digital textbooks. Additionally, the adoption of computers and tablets in education has facilitated increasingly flexible control over the navigation of narrations and videos. Meta-analyses (Ginns, 2005; Reinwein, 2012) and recent studies conducted in primary schools (e.g., Knoop-van Campen et al., 2018, 2019) have demonstrated that pacing control is an influential moderator; these lines of research have not found evidence for the modality effect in learner-paced conditions (see also Tabbers et al., 2001, 2004). However, this moderation might be due to how learner-pacing is implemented in experiments, which differs from classroom applications and modern videos.

Learning time in system-paced conditions has commonly been limited to the length of the narration (Tabbers et al., 2004), whereas learner-paced conditions have had fewer time constraints (Reinwein, 2012), or even unlimited time (e.g., Leahy et al., 2003; Tabbers & de Koeijer, 2010). This can affect the modality effect in two ways. First, the modality effect may be enhanced in system-paced conditions, as slow readers may not have enough time to read the entire text, while an appropriate amount of time for learning is given in the narrated condition (see also Stiller et al., 2009). Second, the modality effect may be reduced in learner-paced conditions, as having more time to process the same amount of information can decrease the demand on working memory (Baddeley, 1992). For example, Knoop-van Campen et al. (2018, 2019) and Witteman and Segers (2010) did not observe the modality effect in the learning benefits of fifth and sixth graders in learner-paced conditions with unlimited time. They also showed that the modality effect did not interact with individual differences in working memory, possibly due to a lower working memory load in conditions with unlimited time (Knoop-van Campen et al., 2019). The present study avoided artificially advantaging either condition by limiting the learning time to approximately two times the length of the narration. This reflects the classroom context, where the learning time is not unlimited, and the time given for an illustrated text is not optimized for its narrated duration.

Another way in which the pacing controls can moderate the modality effect is by enabling the listeners to revisit previous content, thereby making the narration less transient. From this perspective, it is surprising that the provision of pacing controls has diminished the modality effect (Reinwein, 2012), as it could be predicted to benefit especially the narrated conditions. However, the pacing options provided in learner-paced studies have been problematic, and this could be an alternative explanation for the above results. For instance, the narrations in studies of fifth and sixth graders have been played and rewound on a slide-by-slide basis rather than continuously (e.g., Knoop-van Campen et al., 2018, 2019; Segers et al., 2008; Witteman & Segers, 2010). This introduces interruptions because learners must recognize that the previous audio segment has ended to change the slide and manually play the audio of the subsequent slide. Moreover, these actions require the learner to shift their attention from the content to its navigation (Hatsidimitris & Kalyuga, 2013) and to the monitoring of the audio track. For example, Zhu et al. (2020) found in their eye-tracking study that learners in the narrated conditions focused significantly more attention on the navigation buttons because they did not know when the slide's narration would end. In contrast, the narrated condition in the present study utilized YouTube videos, which are the prevalent illustrated narration used in classrooms today. These types of videos can combine benefits from both system- and learner-paced narrations by progressing without interruptions while providing the option for continuous navigation, which may alleviate the transient information effect and reduce the extraneous cognitive load related to the controlling of the presentation.

1.3. The importance of the ecological validity of the settings, materials, and tests in modality effect research

Since previous modality research has mainly studied adults or university students (Reinwein, 2012) in laboratory contexts (Butcher, 2014), more studies of children in authentic classroom contexts are needed (Herrlinger et al., 2017; Mayer & Fiorella, 2014; Schüler et al., 2012; Yang et al., 2013). Valuable steps in this regard have already been taken by Knoop-van Campen et al. (2018, 2019) and Segers et al. (2008), who studied the modality effect in children using materials that were similar to expository texts commonly studied in primary school science lessons in content and length. However, there are still notable differences in settings, materials, and tests between these studies and classroom contexts (i.e., dimensions of ecological validity, see Schmuckler, 2001). As the modality effect has not been shown to be robust to these differences, the generalizability of these results to classroom contexts is unknown. The following provides examples of how the differences between classroom and research contexts can affect the modality effect.

Concerning the settings, previous studies of the modality effect in children have been conducted under varying circumstances: out of the classroom in a one-on-one setting with a researcher (e.g., Knoop-van Campen et al., 2018, 2019), in small groups in the school hallway or auditorium (e.g., Segers et al., 2008; Witteman & Segers, 2010), or different modality conditions in separate classrooms (e.g., Herrlinger et al., 2017). This may influence learning and cognitive load (Choi et al., 2014) in several ways. First, classrooms are specifically designed for learning and thus differ from other settings in physical features such as lighting, noise, and density of people, each of which can affect learning (see Evans, 2006; Higgins et al., 2005). Second, children may pay increased attention to a less familiar environment (Cycowicz, 2019), which would increase extraneous cognitive load. Third, children's prior science knowledge may be context-dependent on their own classroom, where most of their science learning has likely occurred (see Smith & Vela, 2001). Prior knowledge can reduce the modality effect by making the studied information more redundant (Kalyuga, 2014). Fourth, children may perform differently when they are alone with a researcher. For example, they may feel pressure to perform and thus exert similar (and higher) effort in both learning conditions, even if they might have found one of the materials less motivating in a setting with less supervision (see Eysenck & Calvo, 1992). Furthermore, it was recently shown that children's working memory functioning can be lower in the classroom compared to a controlled individual setting (Friso-van den Bos & van de Weijer-Bergsma, 2020), possibly making the reduced cognitive load of narrations more important in a classroom setting. Lastly, testing the modality conditions in separate settings may introduce confounding factors.

Concerning the ecological validity of the materials, especially the pacing controls of the narrated materials used in previous studies have differed from those commonly used in classroom contexts (Section 1.2). In addition, the removal of the written text in studies of the modality effect has left blank spaces in the narrated material. This has been especially notable in studies using longer expository texts, leading to narrated materials that are half-blank (e.g., Knoop-van Campen et al., 2018, 2019; Segers et al., 2008). This suggests that the illustrated text has been the primary basis for the learning material, resulting in narrated materials that would only exist in a research context. Such materials, as well as other unfamiliar situations exemplified above, can affect learning unpredictably because of novelty effects (i.e., a cognitive, affective, or behavioral short-term change due to the initial exposure to a novel situation, see Klein, 2022). For example, the learners may unnecessarily direct their attention to the blank part of the material or wonder if there is an issue with the presentation, which can disadvantage the narrated condition.

In contrast, the learning tests used in prior studies of the modality effect have mostly averted the problems of ecological validity by using familiar types of educational workbook questions, such as multiple-choice retention questions. Additionally, cognitive load has commonly been measured with a self-rating scale, which resembles the self-assessment questions found in modern workbooks. However, some studies have used multiple self-rating scales during learning, which can substantially alter the primary task. For example, in materials used by Tabbers et al. (2001, 2004), every other slide of the presentation was used solely for cognitive load measurement. First, such self-ratings introduce pauses in learning that can be expected to lower cognitive load (see Cheon et al., 2014; Xie & Salvendy, 2000), which could explain why the average of multiple cognitive load measurements during the task has yielded systematically lower cognitive load estimates compared to measurement after the task (see Schmeck et al., 2015; van Gog et al., 2012). Second, stopping the learning process to answer a questionnaire item would likely empty the contents of the learner's working memory, which would hinder their capability to follow along with the presentation and make connections between parts of the content. Third, modality comparisons could be biased because only a continuous video condition requires interaction for pausing. In short, multiple measurements of cognitive load during a presentation would arguably measure the cognitive load experienced in a situation that does not resemble classroom learning, where the learners alternate between studying and answering questionnaire items.

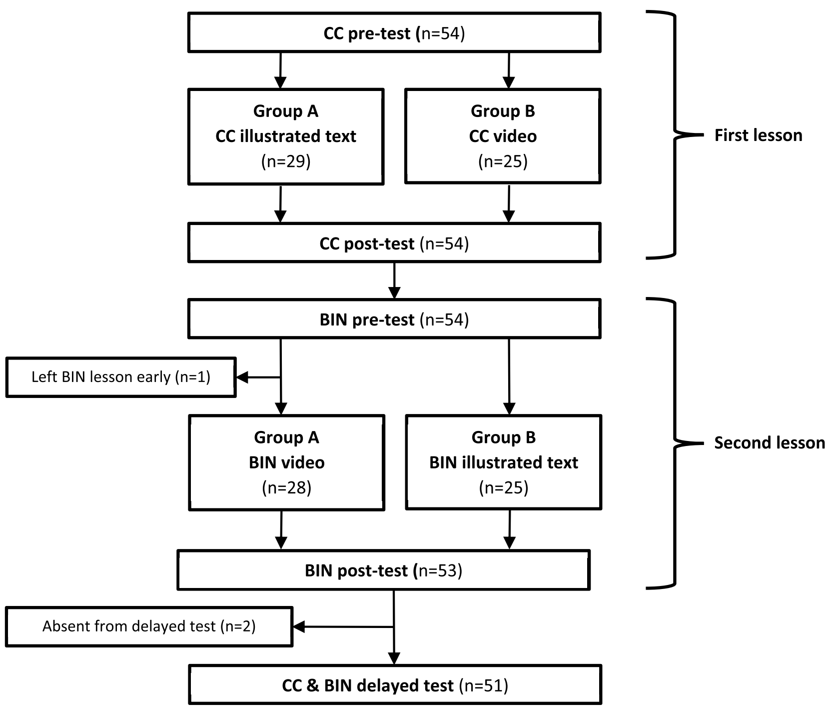


Fig. 1. Study design. CC = climate change, BIN = balance in nature.

2. Material and methods

2.1. Sampling and participants

The data were collected from three fifth-grade classes in a public suburban middle-class primary school in Southwest Finland. The 54 children (26 boys and 28 girls) who participated were 11.9 (SD = 0.4) years of age on average and represented 76% of all fifth graders in the sampled school. Representativeness was further enhanced since approximately 99% of lower primary school students in Finland attend public mainstream schools rather than private schools or special schools (Official Statistics of Finland, 2021).

The research complied with the ethical principles of research with human participants (see Finnish National Board on Research Integrity TENK, 2019). The data were collected anonymously and written informed consent to participate was provided by participating children and their legal guardians.

2.2. Data collection and procedure

The experiment followed a completely confounded factorial design (see Kirk, 2009). In each class, pre-tests, learning interventions, and post-tests were completed in two lessons (Fig. 1), which were instructed by the first author, who is a certified teacher. Children were randomly assigned to either the learning condition group A or group B. During the first lesson, all the children initially completed a pre-test on climate change (5 min), and then group A studied climate change using an illustrated text and group B studied using a video (12 min). After this learning process, all the children completed a post-test on climate change (28 min). The second lesson had the same structure as the first, but the topic was balance in nature, and group A studied using a video while group B studied using an illustrated text. The lessons were 45 min with a 15-min break in between, as customary. Seven days after the lessons, all the children completed a delayed test on climate change and balance in nature.

The children studied and completed the tests in the same environment where they usually study science: working individually in their own classroom and with their class peers present. The children gained access to the learning materials using their personal tablet computers, and the tests were completed on paper. At the time of the experiment, the children had used their personal tablet computers for learning for almost two school years, and thus both learning conditions represented a familiar way of learning for participants. Lastly, the teachers were present to support the children in their individual learning; however, they were instructed to not provide any information related to the topics of study.

2.2.1. Materials

A video and an illustrated text were created for both topics (Fig. 2). These were based on the materials used in Knoop-van Campen et al. (2018, 2019), which were deemed as being similar to materials used in Finnish schools. The materials were translated into Finnish and extended (29% increase in text characters) to better match the available time. Revision of the materials was done in collaboration with a teacher from one of the studied classes to ensure ecological validity, and that the content was appropriately challenging and novel. To represent the variety of images in science textbooks, the materials included both realistic and logical instructional pictures (see Ploetzner, 2012): 12 diagrams, 2 line charts, and 25 photographs.

The illustrated text was presented as a slideshow, in which the text and pictures were displayed side by side. In the video, narration was used instead of written text, and the pictures were scaled and centered to ensure ecological validity. The illustrated texts were accessed via Google Slides: the children could freely change slides forward and backward or skip to any slide. The videos were accessed via YouTube: they progressed without interruptions while providing modern pacing controls (pause, rewind, fast-forward, and continuous navigation using a timeline slider).

The length of the materials reflected that of expository science textbook texts. The climate change materials had 474 words (for

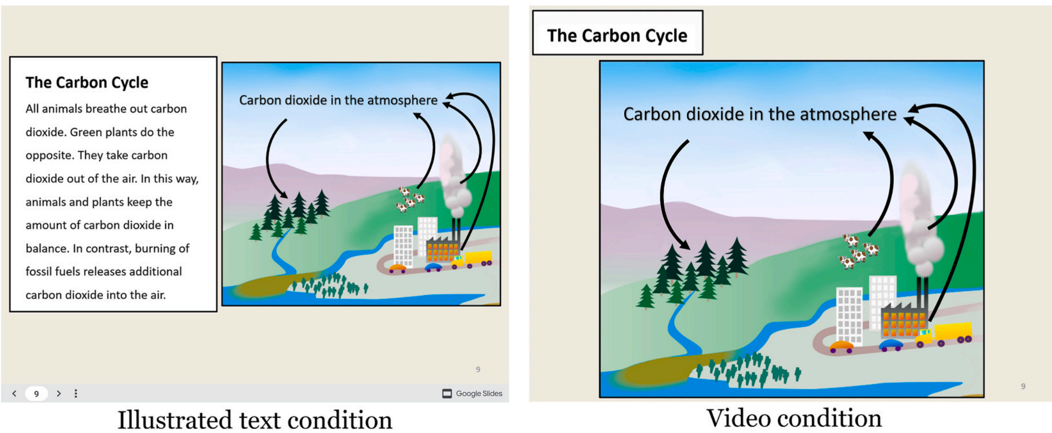


Fig. 2. Examples of illustrated text and video materials (English translation).

reference, 715 in English translation) and 18 pictures on 16 slides, and the video was 5 min and 55 s long, whereas the balance in nature materials had 466 words and 21 pictures on 15 slides, and the video was 5 min and 37 s long. Learning time (12 min) was limited to approximately two times the length of the narration so that the learners in both conditions would have time to finish and review the content. To encourage the children to utilize all the available learning time, all the materials had an additional slide at the end, which listed the topics of previous slides and instructed the children to "go back and review to ensure that you have learned everything."

2.2.2. Measures of domain knowledge

The pre-, post-, and delayed tests were based on the study by Knoop-van Campen et al. (2018). In collaboration with a teacher from a participating class, test questions were revised to reflect the translated learning material and correspond to a style of primary school test that would be familiar to participants. Retention was measured by multiple-choice, pictorial, and closed-ended written answer questions (Appendix A1), which assessed the retention of information included in the material. Transfer was measured by open-ended questions, which required the learner to apply the information to novel contexts. Each question was coded into 1 to 4 items (e.g., question "Name two greenhouse gases" had two items) and each item was scored using points against a predefined rubric (1 point for correct and 0 points for incorrect answers with partial scoring in a subset of items). Item counts can be seen in Table 1. Post-tests and delayed tests had identical retention questions, whereas only a subset of these retention questions was included in the pre-test. Therefore, testing enhanced learning similarly in both learning conditions, although the participants were not aware that the same questions would be used, and the testing did not reveal answers to questions. Transfer was only measured in the post-test.

In within-subject experiments of learning, two separate topics are required to prevent practice effects. However, testing with two tests and materials results in different score means and variations. Therefore, all measures of retention, transfer, and cognitive load were standardized separately for both topics, and the data from both topics were combined before the within-subjects analysis. This procedure enhanced the statistical power and validity, as the aim was to assess how learning would generalize across two science topics, rather than assess differences in learning between the two specific topics. Nonetheless, the scores were similar across topics (Table 1).

2.2.3. Measures of cognitive load and efficiency

Post-tests started with a nine-point mental effort self-rating scale (Paas, 1992) in which the children were asked, immediately after learning, about the level of effort they experienced during the learning process (see van Gog & Paas, 2008) from a *very very low effort* to a *very very high effort*. This measurement technique was chosen because it preserves the authentic classroom context and outcomes as opposed to self-ratings during the presentation, dual-tasks (see Chen et al., 2016, p. 56), and most behavioral measures, such as eye-tracking. Efficiency was calculated as the difference between standardized performance (retention, delayed retention, or transfer) and standardized cognitive load, divided by the square root of two (Paas & Van Merriënboer, 1993).

2.3. Reliability

2.3.1. Scoring

A random 20% subset of the pre-, post, and delayed tests on both topics was independently coded by a second rater. Inter-rater reliability was assessed using two-way mixed, absolute agreement, single-measures intra-class correlations, which were in the excellent range (Table 1; see Koo & Li, 2016). Therefore, minimal measurement error was introduced in the coding process. Differences between the two scorings were discussed and all data were rechecked for discrepancies.

2.3.2. Domain knowledge scale reliability

Post- and delayed test retention and transfer scale Cronbach's alphas were in the range .67–.83 (Table 1), which is considered acceptable in studies of science education (Taber, 2018). Reliability was further enhanced because all analyses used data from the scales of both topics.

Table 1
Descriptive statistics, scale score reliabilities, and inter-rater reliabilities by topic.

Test	Sample size	*N* of items	*α*	*ICC*	Mean	*SD*	Min	Max
CC pre-test	54	14	.16	1.00	19.4	9.14	0.0	35.7
CC post-test retention	54	22	.83	.97	57.9	21.7	4.6	95.5
CC post-test transfer	54	12	.67	.92	37.2	20.0	0.0	75.0
CC delayed test retention	51	22	.78	1.00	51.4	19.7	4.6	90.9
CC learning cognitive load	52	1		1.00	4.71	1.33	2.0	9.0
BIN pre-test	54	11	.26	.95	19.4	10.6	0.0	45.5
BIN post-test retention	53	16	.82	.97	63.6	20.7	6.3	92.2
BIN post-test transfer	53	10	.76	.94	40.5	23.0	0.0	80.0
BIN delayed test retention	49	16	.72	.95	54.4	18.0	4.7	89.1
BIN learning cognitive load	50	1		1.00	4.68	1.60	1.0	9.0

Note. CC = climate change, BIN = balance in nature, α = Cronbach's alpha coefficient, *ICC* = intra-class correlation. Means, minimums, and maximums are percentages of correct answers in the test, except for cognitive load statistics.

Low alphas of the pre-tests ($\alpha = .16$–$.26$) were to be expected due to the students' lack of prior domain knowledge, indicated by low pre-test scores (19.4% correct). Consequently, pre-test scores were not used in the final models. Instead, the inclusion of random effects for participants in a within-subject design controlled for general differences in prior knowledge. Additionally, the random assignment of experimental groups reduced the inter-group differences in prior knowledge between the specific topics of climate change and balance in nature.

2.3.3. *Cognitive load measurement reliability*

The mental effort self-rating scale is widely used (Mutlu-Bayraktar et al., 2019) and is regarded as a reliable and valid estimator of total cognitive load (Chen et al., 2016; Gopher & Braune, 1984; Paas et al., 1994, 2003; Sweller et al., 2019; van Gog & Paas, 2008; cf., Anmarkrud et al., 2019). However, previous studies of the modality effect (e.g., Leahy & Sweller, 2011; Wong et al., 2012) have raised concerns about negative item bias and extreme response bias in children (see Chambers & Johnston, 2002; Marsh, 1986). These issues were avoided because the self-rating scale did not include negative items and extreme responses were rare. Furthermore, Ayres (2006, exp. 2) concluded that subjective ratings provided a reliable estimate of cognitive load in children only two years older than the children in this study. Another criticism relates to subjectivity: participants may understand the rating scale differently (Jiang & Kalyuga, 2020) or rate a similar experience differently. These individual difference factors were controlled by the within-subjects design and the inclusion of random effects for participants in the analysis.

2.4. *Analysis*

Separate linear mixed-effects models (LMMs) were constructed for retention, delayed retention, transfer, cognitive load, and efficiencies. R software (R Core Team, 2021) and R package lme4 (Bates et al., 2015) were used in the modeling. First, random effect intercepts for participants were included in all models to account for individual differences and within-subjects dependencies. Second, the effect of modality (video or illustrated text) was assessed by likelihood ratio tests between models with or without a fixed effect for modality. Third, the effect of cognitive load was assessed similarly in models with a fixed effect for modality. These models are further discussed in section 3. Results.

Since standardized scores were used in the analyses, the reported *B*-values provide effect sizes in terms of standardized mean differences (similar to Cohen's *d*). Furthermore, R package performance (Lüdecke et al., 2021) was used to calculate 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) for each model.

Class-level intra-class correlations were low for all response variables (*ICC*s $< .021$), indicating that teacher effects and other sources of class dependence were sufficiently controlled. Similarly, adding topic (climate change/balance in nature) as a fixed effect to the intercept-only models did not improve the model fit (*p*-values $> .86$), confirming that the effect of the topic was controlled by standardizing the response variables separately for each topic.

Inspections of residual and quantile-quantile plots of the models did not show obvious deviations from assumptions of normality or homoscedasticity. Case deletion diagnostics (Fox, 2020) were applied to each model, which showed that the results were robust to potentially influential observations.

3. Results

Learning occurred across the test phases in an expected manner in both conditions (see Table 2 for descriptive statistics). Before the intervention, the children answered 19.4% (*SE* = 1.0%) of the pre-test questions correctly, which indicates that their prior domain knowledge was low. After studying, the proportion of correct answers in the same questions rose to 59.8% (*SE* = 2.0%) in the post-test, whereas a week later, it regressed to 50.6% (*SE* = 2.0%) in the delayed test. Paired t-tests performed separately for both conditions indicated that both conditions yielded significantly higher scores in post- and delayed tests compared to the pre-tests (*p*-values $< .001$). On average, the children rated their cognitive load during learning as 4.7 (*SE* = 0.1) on a 9-point scale, corresponding most closely to the option *not low nor high effort*.

Table 2
Descriptive statistics of learning, efficiency, and cognitive load across conditions.

	Video condition		Illustrated text condition	
Cognitive load while learning	4.30 (.22)		5.04 (.18)	
	Learning	Efficiency	Learning	Efficiency
Pre-test questions	20.4 (.01)		18.5 (.01)	
Post-test retention questions	64.5 (.02)	.366 (.121)	56.7 (.03)	−.284 (.141)
Delayed test retention questions	57.4 (.03)	.396 (.137)	48.2 (.03)	−.345 (.134)
Post-test transfer questions	41.5 (.03)	.333 (.135)	36.0 (.03)	−.260 (.135)

Note. Values for learning are given as percentages. The numbers in parentheses represent standard errors. One participant was excluded due to absence in the video condition.

3.1. Modality effect in learning

Modality had a significant effect on retention scores, with the video condition outperforming the illustrated text condition in immediate retention, $\chi^2(1) = 10.35$, $p = .001$, $B = 0.345$ ($SE = 0.102$) and delayed retention, $\chi^2(1) = 14.07$, $p < .001$, $B = 0.481$ ($SE = 0.119$). Although the trend was similar, the difference between the conditions on transfer scores was not statistically significant, $\chi^2(1) = 2.82$, $p = .093$, $B = 0.242$ ($SE = 0.142$). The overall models accounted for the majority of the variability in retention ($R^2_{con} = .721$) and delayed retention ($R^2_{con} = .638$). However, the proportion of variance explained by the modality was small ($R^2_{mar} = .030$) and moderate ($R^2_{mar} = .060$) in retention and delayed retention, respectively. These results suggest that while most of the variance in learning performance was explained by individual differences, learning from a video benefited retention, especially delayed retention, in a significant way. However, the variance explained by the transfer model ($R^2_{con} = .451$) suggests that the majority of the variability in transfer scores was due to factors other than individual differences or the modality of the material.

3.2. Modality effect in cognitive load

Cognitive load was significantly lower in the video condition compared to the illustrated text condition, $\chi^2(1) = 12.47$, $p < .001$, $B = 0.527$ ($SE = 0.140$), indicating a modality effect. A large portion of the variability in cognitive load measurements was accounted for in the model ($R^2_{con} = .506$) whereas the modality explained a moderate amount of variance ($R^2_{mar} = .070$).

3.3. The effect of cognitive load on learning

Cognitive load did not have an effect on retention, delayed retention, or transfer when added to the LMMs reported in section 3.1 (p-values $> .457$). Analyzing the data separately for the video condition and the illustrated text condition revealed that the correlations between cognitive load and learning outcomes were low in both conditions ($|\text{correlations}| < .081$). Moreover, inspections of scatter plots did not reveal nonlinear relationships between the cognitive load and learning outcomes in either condition.

3.4. Modality effect in efficiency

The video condition manifested a lower learning cognitive load with higher performance in every learning metric. This led to better efficiency for the video condition compared to the illustrated text condition in retention, $\chi^2(1) = 19.13$, $p < .001$, $B = 0.635$ ($SE = 0.133$), delayed retention $\chi^2(1) = 19.79$, $p < .001$, $B = 0.735$ ($SE = 0.149$), and transfer $\chi^2(1) = 12.54$, $p < .001$, $B = 0.586$ ($SE = 0.157$). These modality effects explained a moderate amount of variance in the efficiency in retention ($R^2_{mar} = .107$), delayed retention ($R^2_{mar} = .139$), and transfer ($R^2_{mar} = .088$).

4. Discussion

The present study analyzed science learning among fifth graders by comparing video and illustrated text conditions in an ecologically valid primary school classroom context. The children studied the topics of climate change and balance in nature individually using their personal tablet computers in a learner-paced setting. According to the results, the video condition outperformed the illustrated text condition in retention, delayed retention, and cognitive load measures. The transfer results followed a similar direction but were not statistically significant. Interestingly, in comparison to the immediate retention test, the modality effect nearly doubled in the one-week delayed retention test. This shows that the relative learning benefit of videos became even stronger over time. The videos were a more efficient way of learning compared to the illustrated texts; they produced better learning outcomes with less effort in the same time frame, suggesting that videos are highly practically effective.

4.1. Explaining the modality effect in terms of cognitive load theory

The present study observed a lower cognitive load for videos compared to illustrated texts. Prior research has produced three explanations for this result. First, the learners can use their working memory more efficiently when receiving both visual and auditory input. Second, the learners do not have to expend cognitive resources on shifting their attention back and forth between the text and the pictures. Third, access to word meanings demands less effort in listening than reading. While all three factors could have contributed to the reduction of cognitive load in the video condition, the last two explanations seem particularly plausible given that (a) a long text similar to one in a science textbook cannot be embedded in a picture, and (b) the participants were children, whose literacy skills were still developing.

Furthermore, cognitive load theory would suggest that this decreased cognitive load accounts for the better learning outcomes observed in the video condition. However, the results were not in line with this, as cognitive load did not explain the differences in learning between the conditions. Additionally, the results in both learning conditions show that the cognitive load was only moderate, and it did not correlate significantly with the learning outcomes. This finding is surprising because the study setting had many qualities that indicate that the reduction of extraneous load could have been a feasible strategy for enhancing learning, such as limited time, novice learners, and relevant pictures (due to the pictorial retention questions). However, there are also factors that might have kept the cognitive load lower in individual learning in the classroom, such as learner-pacing, less transient materials, time for reviewing, and less pressure to perform when compared to a typical research setting. The results suggest that cognitive overload might have been infrequent, which would explain why the reduction of extraneous load was less relevant to the learning benefits of videos.

4.2. Modality effect in a learner-paced study of children

The modality effects observed in the present study contrast with the results of previous studies of children in learner-paced settings, which have mostly produced no modality effects or even reverse effects. One explanation could be that the present study limited learning time to approximately twice the length of the video, whereas the learning time in previous studies was unlimited. However, this explanation is unlikely: despite the unlimited available time in Knoop-van Campen et al. (2018, 2019), their participants' average used learning time is less than that in the present study (compared to the length of the presentation). This highlights the importance of instructions in settings where learners can direct their learning process, and it also might indicate that unlimited learning time undermines motivation. A key difference in the present study is that the children were encouraged to utilize all the remaining time to review the content. Thus, another explanation for the present results could be that the longer learning time may have benefited the video condition by mitigating the transient information effect.

Related to the above, a further explanation for the differences in the results could be the type of learner-pacing employed in the video condition. First, a YouTube video is a more familiar type of learning material for children compared to a slideshow with embedded narrations typically used in previous studies. This may have reduced the modality effect in prior studies, as even beneficial types of pacing controls can at first hinder learning when the learner is adapting to the interface (Hatsidimitris & Kalyuga, 2013). Second, while the video provided the option for learner-control, it progressed without interruptions if the learner did not use the controls. This could make the initial viewing and listening (and hence learning) more time-efficient and reduce the extraneous load related to the controlling of the narrations in previous studies. Third, the video in the present study could be rewound continuously, whereas the narrations in previous studies were segmented into slides. This could have made it easier for the children in this study to locate the desired content and avoid unnecessarily reviewing content that they already knew, potentially benefiting both learning and motivation.

The weaker modality effects in previous learner-paced studies and the stronger effects in system-paced studies can also be explained from the viewpoint of ecological validity. In previous learner-paced studies, using the illustrated text as the primary basis for the materials may have unnaturally disadvantaged the narrated condition in two ways. First, as explained above, the slide-based pacing controls were better suited for written texts. Second, the screen space was optimized for illustrated texts while the narrated materials were half-blank. Conversely, limiting the learning time of the written condition to the length of the narration in system-paced studies is an example of using the narrated condition as the basis of the comparison, potentially leading to the stronger modality effects in system-paced studies because not all learners have enough time to read the entire text. These three applications of learning materials are not commonly used in schools because they artificially remove inherent benefits of different modalities, such as being able to reread written text. Another reason why the present study aimed to ensure that both learning conditions are ecologically valid is that fifth graders can have five years of systematic practice with the types of materials used in schools but no experience with new applications. Therefore, the effect of prior experience could exceed the effect of the modality if just one of the learning conditions were familiar to learners. This could explain why the familiar learner-paced written texts have produced better results compared to the unfamiliar types of learner-paced narrations used in previous studies, whereas the familiar system-paced narrations have outperformed the unfamiliar system-paced written texts.

4.3. Limitations

This study aimed to match the setting, materials, and tests of classroom learning as closely as possible. However, the children were aware that they were participating in research due to ethical reasons, which may have affected their learning process. Additionally, it must be noted that videos may have even more potential that is not reflected in the results of this study, which aimed to match the multimedia content in digital textbooks that are still based on a long tradition of using illustrated texts. Because of this, the materials did not take full advantage of dynamic visualizations and the higher number of pictures per word, which are commonly seen in videos but are hard to implement in illustrated texts. Lastly, the present study aimed to represent the authentic classroom context as a whole but could not identify specific factors in the context that drove the results. Future research could address this by separately assessing distinct aspects of ecological validity and varying them systematically in instructional conditions. However, such an approach should be implemented cautiously: if a resulting learning condition is unfamiliar to learners, the results might reflect novelty effects, such as the lack of practice with an unfamiliar way of learning.

4.4. Individual differences and statistical power

The present study adopted a similar holistic approach to individual differences as it did to the classroom context, by aiming to take into account all individual difference factors while not being able to distinguish specific factors. The results show that individual differences play a key role in the modality effect, accounting for a large portion of variance (44%–69%) in all tests. The within-subjects design enabled the present study to take into account not only the more stable traits like reading comprehension and nonverbal intelligence (which, for comparison, accounted for on average 33% of the variance in similar tests in Herrlinger et al., 2017) but also those factors that remain mostly stable during the two consecutive lessons but can change frequently, such as how well the participants slept the night before or how they interpret a subjective cognitive load item. As it is not feasible to measure all the known and unknown individual difference and contextual factors that may affect learning and cognitive load, within-subjects designs in authentic contexts are recommended in studies of the modality effect, which is sensitive to these factors. Lastly, the within-subjects design enhances statistical power, which is especially important in studies that assess moderators for the modality effect because their goal is to suggest that a missing effect is due to a moderator variable. However, a meta-analysis (Reinwein, 2012) found publication bias in studies of the modality effect and showed that the studies had only 19 participants per between-subjects condition on average. In a comparison between the means of a narrated and a written condition, this sample size facilitates only detecting large effect sizes reliably ($d > 0.93$ is required for 80% statistical power and $p < .05$; see Faul et al., 2007), which means that many prior studies have lacked the statistical power to detect effects like those reported in this paper.

4.5. Conclusions

Despite the widespread use of videos in primary schools, the relative benefits of videos compared to illustrated texts have been unclear due to a lack of research focusing on modern applications of these materials in a classroom context. The results of the present study indicate that videos can be recommended for classroom instruction with respect to learning outcomes, cognitive load, and learning time when compared to illustrated texts. The low class-level variance and the use of two topics suggest that this recommendation is generalizable to similar classes and topics; however, it must be noted that the recommendation applies only to situations similar to the one described in this paper, which nonetheless is a common occurrence in primary schools today.

The results of this study are relevant for practice for four reasons. First, digital textbook providers often give narrated and video options to illustrated texts, which makes the modality effect a key factor for teachers to consider. Second, the rising use of tablet computers in education facilitates a more individualized way of learning from videos, in which children are often provided with pacing controls similar to the present study. Third, teachers do not choose between instructional methods based solely on the learning outcomes, but the learning time and effort must be considered simultaneously to make a realistic decision. Fourth, thanks to the limited learning time, a session of individual learning similar to this study can be easily incorporated into a lesson to build children's topic knowledge, and thereby help them contribute to subsequent activities.

The explanations for the modality effect provided by theoretical frameworks (i.e., cognitive load theory and cognitive theory of multimedia learning, Mayer, 2014, 2020, p. 282–288) are based on the avoidance of cognitive overload. However, these explanations were developed in a system-paced research context. The results of the present study suggest that the avoidance of cognitive overload may not be the primary factor for explaining the modality effect in a typical primary school classroom context, where the learner controls the pace of the presentation. Consequently, the theoretical rationale behind the modality effect in learner-paced settings must be reconsidered.

Funding

This work was supported by the Doctoral Programme on Learning, Teaching and Learning Environments Research (OPPI), University of Turku Graduate School. The funding source was not involved in the decisions made in the research process.

Credit author statement

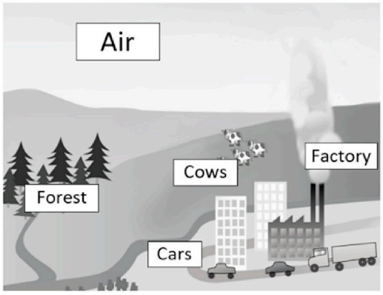
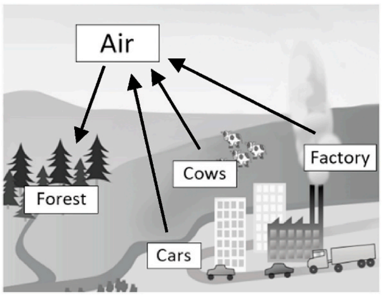
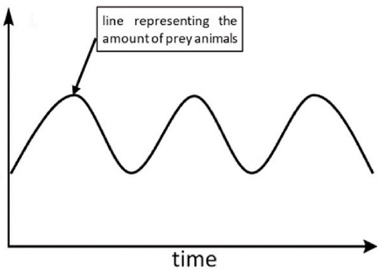
Mikko Haavisto: Conceptualization, Methodology, Formal Analysis, Writing – original draft preparation, Writing- Reviewing and Editing. Tomi Jaakkola: Conceptualization, Methodology, Writing - Reviewing and Editing. Janne Lepola: Conceptualization, Writing - Reviewing and Editing.

Table A.1
Question types and examples (translated to English) of pre-, post-, and delayed tests.

Lesson and question type	Example
Climate change, retention, closed-ended question	Write four examples of fossil fuels. _______, _______, _______, and _______. Correct answer examples: coal, gasoline, diesel, natural gas
Balance in nature, retention, closed-ended question	What type of organisms are on the bottom level of the ecological pyramid? ___________ Correct answer: plants
Climate change, retention, multiple choice	Circle the correct answer. The greenhouse effect is getting stronger because … A. the number of plants has increased. B. the number of animals has decreased. C. of the amount of oxygen in the atmosphere. D. of the use of fossil fuels. Correct answer: D. of the use of fossil fuels.

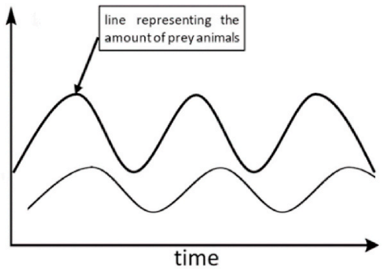
(continued on next page)

Table A.1 (*continued*)

Lesson and question type	Example
Balance in nature, retention, multiple choice	Circle the correct answer. What can the food chain look like in the forest? A. leaf → great tit → larva → hawk B. larva → leaf → great tit → hawk C. larva → leaf → hawk → great tit D. leaf → larva → great tit → hawk Correct answer: D. leaf → larva → great tit → hawk
Climate change, retention, pictorial	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.  Correct answer: 
Balance in nature, retention, pictorial	Draw in the graph a line that shows, how the amount of predators changes with respect to time. Take into account the line of prey animals, which is already drawn.  Correct answer:

(*continued on next page*)

Table A.1 (*continued*)

Lesson and question type	Example
	 *Note.* The drawn line conveys three pieces of information: 1. There are less predators than prey, 2. The number of predators oscillates, and 3. The change in the number of predators follows the change in the number of prey.
Climate change, transfer, open question	Imagine that researchers have invented three machines, which help prevent the climate change. Write below what these machines do and how it helps prevent the climate change. Machine 1:______________________________ ____________________________________ Machine 2:______________________________ ____________________________________ Machine 3:______________________________ ____________________________________ Correct answer examples: 1. The machine produces bicycles so less cars are needed. 2. The machine captures carbon dioxide from the atmosphere. 3. The machine produces meat without using animals.
Balance in nature, transfer, open question	Imagine that you are a minister, whose responsibility is to take good care of nature in Finland. What would you do protect the balance in nature? Come up with three things. A: __________________________________ ____________________________________ B: __________________________________ ____________________________________ C: __________________________________ ____________________________________ Correct answer examples: 1. I would establish more nature reserves. 2. I would declare more species as protected species. 3. I would inform the public about the balance in nature.

Declaration of competing interest

None

Data availability

The authors do not have permission to share data.

Acknowledgements

We would like to thank Carolien Knoop-van Campen and Eliane Segers for contributing the materials, which were used in the development of the tests and materials of this paper.

References

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

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

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

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

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*, Article 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

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

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

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

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.

Fox, J. (2020). *Regression diagnostics* (2nd ed.). SAGE Publications, Inc. https://doi.org/10.4135/9781071878651

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

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

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

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

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

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

Higgins, S., Hall, E., Wall, K., Woolner, P., & Mccaughey, C. (2005). *The impact of school environments: A literature review*. The Centre for Learning and Teaching, School of Education, Communication and Language Science, University of Newcastle. https://www.researchgate.net/publication/232607630_The_Impact_of_School_Environments_A_Literature_Review.

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

Hyvönen, T., Niemi-Sampan, P., & Koli, A. (2018). *Materiaalit ja tekijänoikeudet opetuksessa*. Kuulas Helsinki. https://www.kopiosto.fi/kopiosto/tekijanoikeustietoa/tutkimustoiminta/.

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. *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.

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

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

Klein, J. (2022). Improving the reproducibility of findings by updating research methodology. *Quality and 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

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

Liu, T., Lin, Y., Hsu, C., Hsu, C., & Paas, F. (2021). Learning from animations and computer simulations: Modality and reverse modality effects. *British Journal of Educational Technology, 52*(1), 304–317. https://doi.org/10.1111/bjet.12996

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

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

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.

Merkt, M., Hoppe, A., Bruns, G., Ewerth, R., & Huff, M. (2022). Pushing the button: Why do learners pause online videos? *Computers & Education, 176*, Article 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. https://doi.org/10.1016/j.learninstruc.2011.03.004. S0959475211000247.

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*, Article 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

Official Statistics of Finland. (2021). *Students and qualifications [e-publication]*. Retrieved from https://www.stat.fi/til/opiskt/meta_en.html. (Accessed 20 August 2022) Accessed.

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. (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 & Motor Skills, 79*(1), 419–430. https://doi.org/10.2466/pms.1994.79.1.419

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

Ploetzner, R. (2012). Pictorial representations and learning. In N. M. Seel (Ed.), *Encyclopedia of the sciences of learning* (pp. 2636–2638). Springer US. https://doi.org/10.1007/978-1-4419-1428-6_308.

R Core Team. (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.r-project.org/.

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

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

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

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

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, 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

Stiller, K. D., Freitag, A., Zinnbauer, P., & Freitag, C. (2009). How pacing of multimedia instructions can influence modality effects: A case of superiority of visual texts. *Australasian Journal of Educational Technology, 25*(2). https://doi.org/10.14742/ajet.1149

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

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

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

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

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.