JYX JYVÄSKYLÄN YLIOPISTO UNIVERSITY OF JYVÄSKYLÄ

This is a self-archived version of an original article. This version may differ from the original in pagination and typographic details.

Author(s): Kiili, Carita; Smith, Blaine E.; Räikkönen, Eija; Marttunen, Miika

Title: Students' Interpretations of a Persuasive Multimodal Video About Vaccines

Year: 2021

Version: Published version

Copyright: © 2021 the Authors

Rights: CC BY 4.0

Rights url: https://creativecommons.org/licenses/by/4.0/

Please cite the original version:

Kiili, C., Smith, B. E., Räikkönen, E., & Marttunen, M. (2021). Students' Interpretations of a Persuasive Multimodal Video About Vaccines. Journal of Literacy Research, 53(2), 196-218. https://doi.org/10.1177/1086296x211009296

Check for updates

Original Article

Journal of Literacy Research
1–23
© The Author(s) 2021
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/1086296X211009296
journals.sagepub.com/home/jlr

Students' Interpretations of a Persuasive Multimodal Video About Vaccines

Carita Kiili[1,2], Blaine E. Smith[3], Eija Räikkönen[4], and Miika Marttunen[4]

Abstract

The present study investigated students' (N = 404) interpretations of the main message and use of modes in a persuasive multimodal video on vaccines. It also examined whether students' topic knowledge, language arts grades, and self-identified gender were associated with their interpretations. Students analyzed a YouTube video in which two entertainers demonstrated the importance of vaccinating children. Students' interpretations of the usefulness of vaccines varied in terms of quality of reasoning, which was associated with students' topic knowledge. Notably, many students' interpretations of the use of modes were incomplete, or they did not even mention certain modes in their response. The results suggest that students should be explicitly taught how to interpret different modes and their uses for argumentative purposes.

Keywords

multimodality, comprehension, argumentation, adolescents

Introduction

The rapid integration of the internet into our daily lives has transformed literacy practices in many ways (Leu et al., 2004; Mills, 2010; New London Group, 1996). For youth in particular, multimodal texts, which involve the complex interweaving of visuals, sounds, text, and movement (Jewitt, 2009; Kress, 2010), are becoming a dominating form of communication. On YouTube, people watch more than a billion hours of videos every day (YouTube, n.d.), with young people being the most active consumers

[1]Tampere University, Finland
[2]University of Oslo, Norway
[3]The University of Arizona, Tucson, USA
[4]University of Jyväskylä, Finland

Corresponding Author:
Carita Kiili, Faculty of Education and Culture, Tampere University, P.O. Box 700, 33014 Tampere, Finland.
Email: carita.kiili@tuni.fi

of videos. For example, 94% of young people aged 18 to 24 in the United States use YouTube (A. Smith & Anderson, 2018), acquiring information on an array of topics outside of school (Rideout et al., 2010).

Despite the importance of videos as information sources for today's students, understanding remains limited about their abilities to make sense of digital videos that involve the complex interweaving of multiple modes. Most of the work on interpreting multimodal texts has focused on *static multimodal texts* that combine text, pictures, and graphics (e.g., Barzilai & Eilam, 2018; Cromley et al., 2010; Walsh, 2006). Only a handful of studies have investigated students' comprehension of *dynamic multimodal texts*, including expository narrated animations (Lee & List, 2019; List, 2018) and audiovisual narratives (Kendeou et al., 2008). Salmerón et al. (2020) are among the first to study the effects of persuasive videos among primary school students. They found that compared with textual information, videos had a stronger influence on students' beliefs on a controversial issue. The present study will contribute to this emerging field by examining how adolescents interpreted the main message and use of multiple modes in a persuasive video on a socioscientific topic, and whether their interpretations were associated with their topic knowledge, language arts (LA) grades, and self-identified gender.

Theoretical Approaches

This study examined how students interpreted a dynamic multimodal text, in this instance, a persuasive video on a socioscientific topic. We drew upon three approaches: multimodality, multimodal text processing, and multimodal argumentation. Multimodality, and more specifically, a social semiotics perspective (Kress, 2010), was employed to analyze the video that students watched and to examine how they interpreted the role of different modes in meaning-making. Multimodal text processing (Magliano et al., 2013; Verhoeven & Perfetti, 2008) and multimodal argumentation (Tseronis, 2018) were used to examine students' different interpretations of the main message in the video, especially in terms of the main arguments advanced through the use of multiple modes. Multimodal text processing was also utilized to understand the role of topic knowledge in meaning-making.

Multimodality

This study was guided by a social semiotics (Kress, 2010) view of multimodality that emphasizes how various modes—including, but not limited to, visuals, sounds, text, motions, and gestures—are integral in meaning-making. In this approach, language is no longer privileged in communication, and other modes have the potential to contribute equally to meaning (Jewitt, 2009).

A social semiotics perspective thus explores how meaning is created through synergistic interactions between different modes. The unique interweaving of different modes communicates messages that no single mode can express independently (Jewitt, 2009; Stein, 2008). Hence, endless combinations and intersemiotic relationships can be created by combining different modes (Unsworth, 2008). For example, co-occurring modes can

align to emphasize a complementary message (Dalton et al., 2015) or diverge to create dissonance and convey different messages simultaneously (Unsworth, 2008). In addition to offering endless communicative possibilities for composers, these multidimensional messages also open up multiple interpretations for audiences (Serafini, 2010). Viewers of multimodal texts attune to different modalities and are tasked to make inferences between them (Walsh, 2006). Furthermore, interpretations of multimodal designs can differ based on modal density and complexity (Norris, 2009).

Central to social semiotics is the understanding that modes comprise semiotic resources and unique affordances for communication. These modal affordances—based on their social histories, cultural uses, and material features—offer potentials that render it more suitable than other modes for certain communicative tasks (Kress, 2010). Some multimodality research has examined how composers leverage the semiotic resources of specific modes (e.g., visuals, sound, and video) in their digital products (Hull & Nelson, 2005; Shanahan, 2013), as well as how modal affordances vary across composers, genres, and intended messages (B. E. Smith et al., 2016). Although most multimodal research has focused on how adolescents orchestrate multiple modes in digital environments to achieve a variety of compositional goals ranging from identity expression to sharing their voice in empowering ways (Hull et al., 2010; Mills, 2010; B. E. Smith et al., 2021), we are not aware of any research that has focused on how adolescents identify and analyze specific modal affordances in dynamic multimodal texts, such as videos.

Processes of Interpreting Multimodal Texts

Several researchers have theorized the processes involved in comprehending multimodal texts (Magliano et al., 2013; Verhoeven & Perfetti, 2008). The multimodal text-processing model by Verhoeven and Perfetti (2008) expands the construction–integration framework (Kintsch, 1998) to understand the comprehension of texts that contain pictorial and textual information. The multimodal text-processing model is based on the idea of separate processing channels for verbal and visual information (Paivio, 1990). This has led to the construction of two separate models: textual and pictorial models. The text model is constructed by identifying words and processing sentences, whereas the pictorial model is constructed by processing visual information. According to Verhoeven and Perfetti (2008), integrating information from the text and pictorial models with the reader's prior knowledge allows the reader to construct a situational model—a representation of the situation presented in the text, including events, actions, and people and in general an understanding of what the text is about (Kintsch, 1998).

Furthermore, the work on processing visually based narratives by Magliano et al. (2013) explored the similarities and differences between processing text- and visually based narratives. They differentiated front-end and back-end processes of constructing an understanding of narrative media, including the processing of language and visual information. Front-end processes are involved in information extraction and encoding, whereas back-end processes are involved in building a situational model. Front-end

processes related to the text- and visually based narratives are different. The former involves orthographic, phonological, lexical syntactic, and semantic processes, whereas the latter involves gist processing. Gist processing refers to extracting global information about the scene (e.g., a café), key objects (e.g., a counter), and actions (e.g., paying). This information guides viewers to gather more information to build an event model.

While front-end processes vary across presentation formats, back-end processes seem to be more consistent across the medium. Such back-end processes include event segmentation, inferencing, and structure building and result in a situational model—that is, a representation of the situation presented in the text or in the visual narrative. Both of the aforementioned approaches, multimodal text-processing and processing visually based narratives, specify inferential processes and background knowledge as integral components when building a coherent representation of multimodal texts, an issue also highlighted in reading comprehension of monomodal texts (Kintsch, 1998; McNamara & Kintsch, 1996).

Most of the reading research on processing videos has focused on examining the acquisition of procedural knowledge through video tutorials (e.g., Arguel & Jamet, 2009) or declarative knowledge delivered by instructional videos (e.g., Merkt et al., 2011) or animations (e.g., Lee & List, 2019; List, 2018). A recent study by List (2018) compared how undergraduates comprehended and integrated information from two texts that were presented via either texts or animated videos. No differences in comprehension or integration performance were found. However, the materials, designed for educational purposes, were straightforward and easily comprehensible expository texts for undergraduates. In the present study, students interpreted a complex, multimodal video that was rich in its use of different modes.

Multimodal Argumentation

Argumentative reasoning, constructing, evaluating, and comprehending arguments, is a foundational aspect of human thinking and learning (Alexander, 2016; Iordanou et al., 2016). In argumentative reasoning, "argument" refers to a claim or an opinion and one or more supportive reasons that are connected with a warrant (Toulmin, 1958). Reznitskaya et al. (2007) have proposed a basic argument schema that is particularly suitable for teaching and learning argumentation in basic education. The proposed argument schema highlights the most crucial elements of argumentation, including position, reasons for the position, supporting facts as well as an objection, and response to it. To comprehend an argumentative text, a reader needs not only to identify the claim and reasons for and against it but also to be aware of the author's intention, the author's position on the issue, the context of the argumentation, and the reader's own understanding of the issue (Haria & Midgette, 2014). For example, in persuasive texts, authors' intention is to change readers' beliefs on a controversial issue (Iordanou et al., 2016).

Although the verbal mode has been dominant in argumentation research, multimodal argumentation has begun to interest researchers (Blair, 2015; Howell et al., 2017; Kjeldsen, 2015). In multimodal argumentation, nonverbal and verbal modes can

contribute directly or indirectly to conveying an argument (Tseronis, 2018). Because argumentative messages may include elements that bear symbolic or intertextual meanings, readers play an active role in reconstructing the argument by interpreting verbal and nonverbal cues (Kjeldsen, 2016). For example, in addition to their content, visuals can contribute to argumentation through color, framing, arrangement, or form (Tseronis, 2018). Students' skills in analyzing and producing argumentative monomodal texts have been extensively studied over recent decades (for a review, see Newell et al., 2011). However, little is known about students' abilities to make sense of multimodal argumentation using a rich repertoire of modes.

Present Study

This study examined students' interpretations of a persuasive video in the context of a larger online inquiry task on a socioscientific topic, namely, vaccines. Students' interpretations were examined from two angles: interpretations of the main message and interpretations of the use of different modes in communicating the main message. Because prior knowledge plays an important role in building a coherent representation of multimodal texts (Magliano et al., 2013; Verhoeven & Perfetti, 2008), we investigated the role of topic knowledge in students' interpretations. Furthermore, as previous research has shown that girls outperform boys in reading comprehension of both printed (Organisation for Economic Co-operation and Development, 2013) and digital texts (Salmerón et al., 2018), we examined for gender differences in students' interpretations of a persuasive video. Finally, we examined students' mean grade in LA as an indicator of academic success in literacies. The specific research questions were the following:

Research Question 1: How did students interpret the main message and supporting reasons in a persuasive video on vaccinations?
Research Question 2: Which modes did students identify as the most important in communicating the main message of the persuasive video, and how did they interpret the use of these modes in communicating the main message?
Research Question 3: Are students' topic knowledge, mean grade in LA, and gender associated with their interpretations of (a) the main message and (b) the use of different modes?

Method

Participants

Participants were 404 students from seven upper secondary schools in Finland. The mean age of students was 17.39 ($SD = 0.41$) years; 243 were females (60.15%), 158 were males (39.11%), and three students abstained from assigning themselves according to the traditional gender binary. Written consent was requested from the students and, if a student was underaged, also from their guardian(s).

Individual Difference Variables

Topic knowledge. To assess students' topic knowledge, we presented them with 10 statements on vaccinations, of which three were correct and seven false. To ensure a reasonable degree of content validity, two university lecturers in health sciences and a medical expert reviewed the statements. Based on the experts' comments, four statements were modified. Students were asked to select three statements that they believed to be correct. Four statements (e.g., Vaccines cause autism) did not differentiate students well as more than 90% of students responded correctly. These items were omitted from the present analyses.

Each remaining item (six items) was scored 0/1. One point was given from each selected correct statement (e.g., A vaccine may contain weakened pathogens) and each nonselected incorrect statement (e.g., It is unnecessary to vaccinate against milder diseases [e.g., chicken pox]). The maximum score was six points. Because the measure included binary items, the score reliability was estimated using a latent variable modeling approach (Raykov et al., 2010). The reliability was .78 (95% confidence interval [CI] = [.72, .83].

LA grades. Students' grades from LA courses were obtained from schools for those who gave permission (N = 383). For the analyses, the mean of the language course grades was used.

Task

The multimodal task was a part of a larger online inquiry task that students completed in a web-based environment in their classrooms. In the web-based environment, students were presented with a task scenario that asked them to explore the benefits and disadvantages of vaccinations. The goal of this exploration was to compose a letter to inform Finnish members of parliament on whether the vaccination of children should be obligatory. This study examined students' interpretations of a persuasive video on the topic.

In the video analysis subtask, students were given a link to the video that opened in a pop-up window. On the main window, there was a form with questions about the video. Students were first asked to answer the following question: What is the main message of the video, and what are the most important reasons supporting this message? The second question was multiple choice and asked students to select the three most important modes used in communicating the main message of the video from five options (audio, motions, text, visuals, and gestures and facial expressions). Examples of each of the five modes were given in parentheses to ensure that students had as similar an understanding as possible on the different elements composing each mode. Finally, students were asked to describe how the three modes they selected were used in the video to communicate its main message.

Students worked on a computer equipped with headphones for listening to the video. They could watch the video as many times as they wished. The questions on the video were presented on the form embedded in the online assessment. To progress in the assessment, students were required to answer all the questions on the video. However, students were able to go back to revise their responses while completing the task within the given time frame.

Video

In the YouTube video used in this study, Penn and Teller, two entertainers who combine magic and comedy, demonstrated to the viewer how vaccines protect children from serious diseases, the aim being to persuade all parents to vaccinate their children (UltraMiraculous, 2010). The video, which has more than six million views, was a segment from the *Penn & Teller: Bullshit!* television show that aired in 2010.

The video begins with the two men walking toward the camera while introducing themselves and the topic: "Hi, I'm Penn and this is my partner Teller. You may have heard that vaccinations cause autism in 1 out of 110 children" (see Online Appendix B for the multimodal transcript). Next, the entertainers strongly refute the statement using colorful language and impassioned facial expressions and gestures. They set up an imaginative experiment to illustrate their position with two groups of children represented by brightly colored bowling pins. One group, representing vaccinated children, was protected by plexiglass, whereas the other group had no such protective barrier. While Penn provides statistics on how vaccines have protected children in the past from diseases like diphtheria and polio, both entertainers hurl yellow balls that represent a variety of diseases (mumps, polio, chicken pox, etc.) at each of the two groups of children. After throwing the balls, the group of children behind the plexiglass shield is still standing, whereas many children in the unvaccinated group have toppled over, or "died." Penn concludes, "We have vaccinations against all of them [referring to the diseases listed]; which side do you want your child to stand on?"

The 91-s-long video includes a variety of multimodal means of persuasion—ranging from the content and delivery of the entertainers' speech through music and sound effects to visual demonstrations through gestures, props, and motions—of the negative effects of not vaccinating children. To identify the persuasive message and supporting arguments, students needed to make inferences based on the different modes used.

We decided on using this video, although it is in English, because it is rich in the use of modes, and we were not able to find anything similar in Finnish. As there are only about five million people speaking Finnish as the first language, the internet offers far fewer videos in Finnish than in English. The video was voiced in English with Finnish subtitles, although many students would have understood the English on the video quite well, having studied it for 8 years in school. In Finland, all TV programs that are not in Finnish are subtitled. Thus, reading subtitles is a daily literacy practice for Finnish students.

Data Analysis

Students' responses on the open question concerning the main message were, on average, 32.90 ($SD = 16.63$) words long. The mean length of the responses concerning the use of the three most important modes was 44.44 words ($SD = 33.76$). Because of the nature of the Finnish language, the responses would have been longer in English.

The analysis of open responses on students' interpretations of the main message and how modes were used to make meaning followed a two-step procedure. In the first step, we applied qualitative analysis to create a coding scheme by using 20% of the data. To do this, we used 81 first responses in the data sheet, which was organized according to

the students' ID numbers. Both coding processes were iterative and involved several rounds of collaborative coding (Strauss & Corbin, 1998). First, we openly coded a sample of the students' responses to generate emergent categories for answering the research questions on how students interpreted the main message and how they connected different modal elements to meaning-making. Throughout the analytic process, we carefully compared students' responses against the multimodal transcript of the video (see Online Appendix B) and identified similarities and differences in students' responses. Next, using axial coding, we traversed the student responses to relate and develop definitions of codes. This process required several rounds with the researchers discussing and refining categories. This qualitative analysis resulted in the coding schemes (see Tables 1 and 2). In the second step, the remaining data were coded by using the developed coding schemes. The interrater reliability was assessed for 20% of the students' responses. Interrater reliability is reported below along with the description of each analysis.

Student interpretations of the main message in the persuasive video. Students' interpretations of the main message of the video and supporting reasons were qualitatively analyzed from two perspectives: (a) usefulness of vaccines and (b) safety of vaccines. The patterns found on the usefulness of vaccinations are presented in Table 1 and show that students' reasoning varied in its sophistication. At Levels 4 and 3, students express one way or another that vaccines are useful. However, at the highest level of reasoning (Level 4), students explicitly connect the piece of evidence to support the usefulness of vaccines, whereas at Level 3, students do not make the connection explicit. To illustrate this difference, one can compare the responses at Level 4, Student 1019, and Level 3, Student 1115 (see Table 1). Student 1019 explicitly states that before the vaccines, many children died from certain diseases, whereas Student 1115 points out the evidence, that is, the numbers from the 1920s and today, without explicating what the numbers tell. At Level 2, students' reasoning is incomplete as it does not include both components of the argument, main message, and supporting reason. Finally, at Level 1, students' reasoning is overgeneralized.

Table 1. Levels of Reasoning in Students' Interpretations of the Main Message of the Video on the Usefulness of Vaccines.

Category	Description	Example response
Level 4: Relevant reasoning for usefulness	Student states that children should be vaccinated OR that vaccines are useful and presents at least one piece of relevant evidence.	It is irresponsible to not vaccinate one's children. The most important justifications are that before the invention of vaccines many children and adults died of certain diseases. Owing to of vaccines, many of these diseases have disappeared. (Student 1019) The video is for vaccines and encourages people to vaccinate their children . . . Unvaccinated children are at multiple risk for falling ill compared to those who are vaccinated. (Student 1018)

(continued)

Table 1. (continued)

Category	Description	Example response
Level 3: Underdeveloped reasoning for usefulness	Student states that children should be vaccinated OR that vaccines are useful but presents evidence that is underdeveloped: (a) evidence is imprecise, (b) only source(s) of evidence are mentioned, (c) the demonstration in the video is described without inferences being made about it.	(a) Video compares the number of people who fell sick in the 1920s to the corresponding number these days. (Student 1115) (b) The claim [that vaccines protect from diseases] is justified with statistical knowledge. (Student 3047) (c) Vaccination is good for children . . . When Penn and Teller throw balls representing diseases, the plexiglass representing vaccines stops every ball whereas a large number of children on the unvaccinated side fall down. (Student 2063)
Level 2: Incomplete reasoning for usefulness	Student states usefulness of vaccines but does not present any relevant evidence for the statement OR provides evidence without presenting an explicit statement on usefulness.	Vaccines protect from diseases and death caused by diseases. (Student 3016) The prevalence of diseases has decreased because of the use of vaccines. (Student 2025)
Level 1: Overgeneralized reasoning for usefulness	Student overgeneralizes evidence: All unvaccinated children will die; vaccines provide full protection.	If you vaccinate your child, she or he does not get any diseases. (Student 2032) If you do not vaccinate your child, she or he will die. (Student 6016)
Level 0: No mention of usefulness		

On the safety of vaccines, students tended to produce interpretations that were either positive (e.g., vaccines are safe, vaccines are not harmful, the risks are minimal or do not cause autism) or negative (e.g., vaccines cause or may cause autism). Based on these analyses, two variables were formed: interpretation of usefulness and interpretation of safety. After identifying the patterns and defining the levels of reasoning, the first author and a research assistant scored 20% of the students' responses. The Kappa value for the usefulness of vaccines was .83 and for the safety of vaccines .88.

Students' interpretations of the use of modes in communicating the main message. To analyze students' interpretations of how the modes (i.e., audio, motions, visuals, text, gestures, and facial expressions) were used to communicate the main message in the video, we first identified sections in students' responses that were related to each self-selected mode. Each of the selected modes was then coded into one of five categories (see Table 2)

according to the highest level of interpretation that each student's response represented. Connecting the selected mode to the argument(s) in the video was regarded as the most sophisticated level, owing to the persuasive purpose of the video.

Based on these analyses, four variables were formed: interpretation of visuals, interpretation of audio, interpretation of motion, and interpretation of gestures and facial expressions. The textual mode was not included in the analysis because most of the students who selected this mode (n = 30) analyzed audio. After identifying the patterns in interpretations of modes, the first and second authors scored 20% of the students' responses, reaching Kappa value of .774.

Table 2. Students' Interpretations of the Use of Modes in Communicating the Main Message of the Video.

Category	Description	Example response
Selected mode is connected to main argument(s)	Student identifies at least one specific multimodal element and connects it to the main argument(s) of the video.	On the other hand, the transparent "protective wall" illustrates the safety of and need for vaccination. (Visual, Student 1013)
Selected mode is connected to the topic	Student identifies at least one specific multimodal element and connects it to the topic of the video.	The person in favor of vaccinations speaks using humor to his advantage. The person opposed to vaccinations does not speak at all. This helps to create an image of uncertain people who are afraid of vaccinations. (Audio, Student 2002)
Selected mode is connected to communicative purposes	Student identifies at least one specific multimodal element and describes how it was used to convince the audience or capture their attention without connecting it to the topic of the video.	The video uses informal speech, which also makes it more easily understandable. This would not be the case if it included challenging medical language. The speech in the video is carefully chosen for the target audience to help them to understand. (Audio, Student 1138)
Incomplete interpretation of selected mode	(a) Student only identifies at least one specific multimodal element.	The men are throwing balls at pins depicting children. (Motion, Student 2025)
	(b) Student does not specify a specific multimodal element but connects the mode in general to the main arguments, overall topic, or ways of convincing the audience.	Auditory elements in the video were supposed to be funny to get people to realize the benefits of vaccines. (Audio, Student 6008)
Selected mode is not mentioned		

Statistical Analyses

All statistical analyses were conducted using Mplus 8.0 software (Muthén & Muthén, 2017). The method of estimation in all analyses was the maximum likelihood estimator with robust standard errors (MLR). Missingness in the data was handled by using the full information maximum likelihood (FIML) procedure (Enders, 2010). FIML uses all the available information in the data to estimate the model without imputing the missing values. As the data were hierarchical in nature (i.e., students were nested within seven schools), unbiased standard errors of the parameter estimates were estimated using the COMPLEX option in Mplus.

Associations of students' topic knowledge, mean grades for LA, and gender with students' interpretations of the main message of the video about vaccines (Research Question 3a) and interpretations of the use of different modes in communicating the main message (Research Question 3b) were examined using multinomial logistic regression (MNL) analysis. Each dependent variable was analyzed separately. Parents' education was used as a control variable.

Furthermore, to compare each interpretation category with all the other interpretation categories, each interpretation category (here called category of interest) was used in turn as a reference category in all MNL analyses. In the "Results" section (See Online Appendixes C–G), MNL regression coefficients are presented as odds ratios (ORs). The interpretation of ORs differs slightly for categorical and continuous independent variables. For categorical independent variables (e.g., gender: 0 = boys, 1 = girls), an OR greater than 1 indicates that students for whom the value of the independent variable is 1 are more likely to be in the interpretation category of interest than students for whom the value of the independent variable is 0. An OR smaller than 1 indicates the opposite. For a continuous independent variable (e.g., mean of LA grade), an OR greater than 1 indicates that the higher the value of the independent variable, the greater the likelihood that the student's response represents the interpretation category of interest. The interpretation of an OR smaller than 1 is the opposite. The statistical significance of ORs was determined by computing 95% CIs for each OR. An OR was statistically significant if the 95% CI did not include the value 1.

Results

Descriptive Statistics

Descriptive statistics for all the measured variables are presented in Online Appendix A. In the following three sections, we first describe the distributions of the dependent variables and then report the results on the associations of individual difference variables with the students' interpretations of the main message of the video and the use of modes.

Interpretations of the Main Message of the Video

Most students (93.6%) mentioned the usefulness of vaccines when interpreting the main message of the video. However, only a quarter (23.8%) of the students provided relevant

reasoning for the usefulness (see Table 1 for examples). The remainder either provided underdeveloped reasoning (22.3%), incomplete reasoning (39.9%), or overgeneralized reasoning (7.7%) or did not mention usefulness at all in their response (6.4%).

The safety of vaccines was less often mentioned (58.2%) than their usefulness. Interestingly, conflicting interpretations were made on safety. While 38.6% of students interpreted vaccines as safe or not a cause of autism, 19.6% of students interpreted the video as presenting the idea that vaccines cause or may cause autism. The video contained two somewhat contradictory messages on the relationship between vaccines and autism that probably confused some students, although only five students had shown a misconception on autism (i.e., had selected the statement that vaccines cause autism in the topic knowledge test) before watching the video. The following examples illustrate how differently two students interpreted the statements related to autism:

> The main idea of the video is to defend vaccines and support the fact that vaccines do not cause autism, even though the antivaccination movement argues so. (Student 3081)

> The main message is that even though one child out of 110 children gets autism because of the vaccines, vaccines protect children from other fatal diseases. (Student 2015)

Identification of the Most Important Modes

Of the five modes, 93% of the students included audio, 91% visual, 69% motion, and 40% gestures and facial expressions among the three most important modes for meaning-making in the video. Only 30 students (7%) selected the textual mode, and many of these were analyses of speech, which in the video context was considered audio.

Interpretations of Use of Modes in Communicating the Main Message

Figure 1 illustrates the distributions of students' interpretations of how different modes were used to communicate the main message. Notably, many students' interpretations were incomplete, varying from 26% to 41% across modes. In addition, many students did not mention the selected mode at all in their responses. The proportion of these students varied widely across modes, from 8% to 29%.

For each of the four modes, the proportion of students who connected it to the main argument(s) of the video varied from 6% to 30%. When students' performance was examined across modes, 35% connected a specific multimodal element(s) to the main arguments in at least one of their interpretations of the three self-selected modes. Of the students who chose the visual mode as one of the three most important, 30% were able to connect a specific visual element to the main argument(s) of the video. Of the students who chose to interpret motion, 17% made a connection between the specific multimodal element and the main argument(s). It seems that students found it more difficult to connect the remaining two modes, audio and gestures and facial expressions, to the main argument(s) of the video. When students interpreted audio, they were more attentive to the ways in which sounds were used for communicative purposes (e.g., how music set a tone or how Penn emphasized certain words) than to

how audio was used in building argument(s) or describing vaccinations. This was somewhat unexpected as, for example, some evidence for the usefulness of vaccines was presented directly in speech.

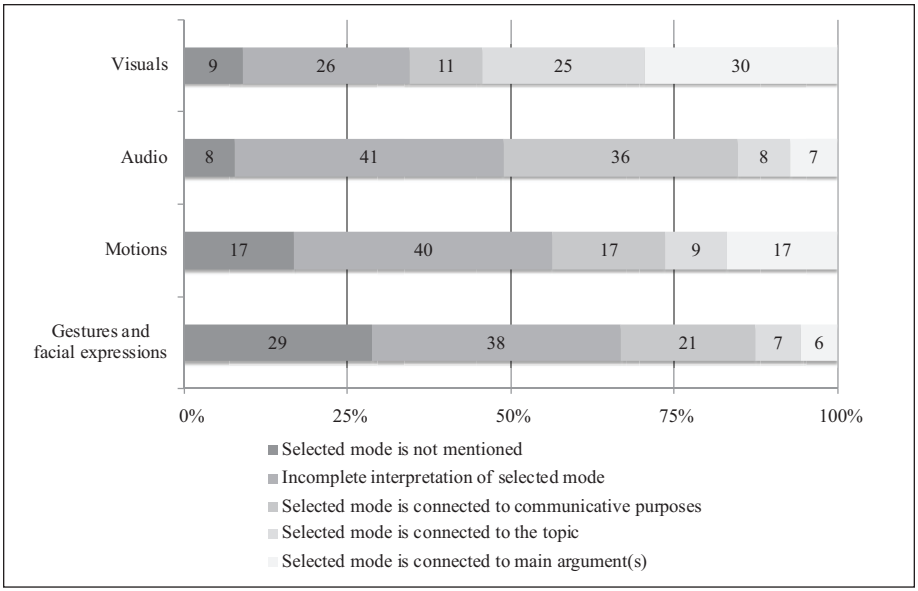

Figure 1. Students' interpretations of the use of modes in communicating the main message of the video.

Associations Between Individual Difference Variables and Interpretation of the Main Message

Results on the associations of individual difference variables with the interpretation of the main message showed that prior topic knowledge was the most consistent predictor of the level of reasoning on the *usefulness* of vaccines. The better a student's prior knowledge on the topic was, the more likely he or she was to provide relevant reasoning than a less sophisticated level of reasoning on the usefulness (see Online Appendix C). In addition, students' mean grade in LA differentiated some levels of reasoning. The higher a student's mean grade in LA was, the more likely he or she was to provide relevant, underdeveloped, or incomplete reasoning than overgeneralized reasoning on the usefulness of vaccines (see Online Appendix C). Self-identified gender was not associated with the level of reasoning on the usefulness of vaccines; thus, the level of reasoning on the usefulness of vaccines was similar among girls and boys.

Students' topic knowledge was associated with the level of reasoning about the *safety* of vaccines. The better a student's topic knowledge was, the more likely he or

she was to give a negative than a positive interpretation of the safety of vaccines (OR = 1.22, CI = [1.08, 1.38]). The other individual difference variables did not differentiate students on their level of reasoning on safety.

Associations Between Individual Difference Variables and Interpretations of the Use of Modes

Results on the associations of individual difference variables with the interpretation of the use of the modes showed that mean grade in LA and topic knowledge were associated with students' interpretations of how *visuals* (n = 275) were used in communicating the main message of the video (see Online Appendix D). The higher a student's mean grade in LA was, the more likely he or she was to connect visuals to the main argument(s) of the video, topic, or communicative purposes than to give an incomplete interpretation of the use of visuals or not to mention the visuals at all.

Topic knowledge differentiated only the students representing the two least sophisticated categories of interpretations: The better a student's topic knowledge was, the more likely he or she did not mention visuals at all in his or her response compared to providing an incomplete interpretation. Students' gender did not predict their interpretations on the use of visuals.

Gender, mean grade in LA, and topic knowledge were associated with students' interpretations on how *audio* was used in communicating the main message of the video (see Online Appendix E). Among the students who chose audio (n = 271) as one of the most important modes, boys were more likely not to mention audio at all in their response than girls, whereas girls were more likely to connect audio to communicative purposes or to provide incomplete interpretation than boys. The higher a student's mean grade in LA, the more likely he or she was to interpret audio than not to mention auditory elements at all in his or her response. Finally, topic knowledge differentiated only the students representing the two least sophisticated categories of interpretations: The better a student's topic knowledge was, the more likely he or she did not mention auditory elements at all in his or her response compared to providing an incomplete interpretation. Thus, topic knowledge was at a similar level for those students whose responses represented the three highest-level categories.

On the issue of students' interpretations of the use of *motion* (n= 209), mean grade in LA was the most consistent predictor (see Online Appendix F). The higher a student's mean grade in LA, the more likely he or she was to interpret motions than not to mention motion elements at all in his or her response. In addition, we found two associations related to topic knowledge. The better the student's topic knowledge, the more likely he or she was to connect motions to the main arguments of the video than to the topic of the video or communicative purposes. Gender did not predict students' interpretations of the use of motions.

All individual difference variables predicted students' interpretations of the use of *gestures and facial expressions* (n = 117) in communicating the main message of the video (see Online Appendix G). Boys were more likely than girls to connect gestures

and facial expressions to the main arguments than to connect them to the topic or to communicative purposes or not to mention them at all.

Mean grade in LA predicted students' interpretations as follows: The higher a student's mean grade in LA was, the more likely he or she was to interpret gestures and facial expressions than not to mention motion elements at all in his or her response. Moreover, the higher a student's mean grade in LA was, the more likely he or she was to give an incomplete interpretation than to connect gestures and facial expressions to the main argument(s) of the video. Finally, the better a student's topic knowledge was, the more likely he or she was to connect gestures and facial expression to the topic of the video than to make other types of interpretations or not to mention gestures and facial expressions at all.

Discussion

This study examined how upper secondary school students interpreted a persuasive video—a type of dynamic multimodal text that they may well encounter when exploring socioscientific issues online. This study contributes to literacy research in several ways. First, according to our knowledge, it is one of the first studies to examine adolescents' interpretations of a persuasive video that is rich in its use of different modes. Second, while most studies investigating multimodal literacies have been small-scale qualitative studies (see B. E. Smith et al., 2021, for a review), the present study included about 400 students. This large sample enabled us to investigate differences between individuals in their topic knowledge on vaccinations, academic success in LA, and self-identified gender in relation to their interpretations of both the main message and the use of different modes in a persuasive video. Finally, the coding scheme developed for students' interpretations of multimodal meaning-making can be applied and built upon in future studies as well as used to inform instruction.

Interpretation of Main Message Communicated in the Video

We examined students' interpretations of the main message of a persuasive video on the importance of vaccinations. Specifically, we investigated whether they were able to identify and interpret the two main arguments supporting children's vaccination. While most of the students identified the argument on the usefulness of vaccines, we found considerable variation in the sophistication of their reasoning. Only a quarter of the students provided relevant reasoning for usefulness; the reasoning of the others was underdeveloped, incomplete, or overgeneralized. For example, some students' interpretations lacked inferences that would have bridged the demonstration in the video to precise arguments. When considering these results in the light of processing of multimodal texts, it might have been challenging for some students to build a coherent situational model when they were required to integrate verbal and visual information with their prior knowledge (Verhoeven & Perfetti, 2008).

Interestingly, the students (58%) who mentioned the safety of vaccines as a main message of the video provided conflicting interpretations of safety. While some students

interpreted vaccines as safe or not a cause of autism, others interpreted vaccines as a cause or possible cause of autism. This might be explained by the multidimensional nature of multimodal texts (Kress & Van Leeuwen, 1996; Norris, 2009; Serafini, 2010). Different modalities contribute different information and can combine in ways that may convey conflicting messages (Unsworth, 2008). Furthermore, different students may attune to different modalities when viewing a video and interpreting its content. However, in the absence of online process data, we cannot offer any stronger explanation for these contradictory interpretations.

Interpretations of Use of Modes in Communicating the Main Message

In addition, we examined students' interpretations of how different modes were used to communicate the main message of the video. Compared with other modes, students most frequently connected visuals to the argumentation of the video. A plausible reason for this is that the visual demonstration of vaccines protecting children from serious diseases was a prominent element in the video. As motions in the video were closely related to the visual demonstration, they were also more frequently connected to argumentation than the other modes. Although many arguments were also presented through audio (i.e., speech), students very seldom connected audio to the argumentation of the video. Instead, they paid more attention to how the audio was used to attract audience attention. As the present students' analyses are highly dependent on how the video utilized different modes, one should be hesitant about drawing conclusions on these students' general abilities to interpret different modes. More studies are needed that employ a diverse set of videos and a more controlled research design to better understand students' interpretation skills across modes.

Individual Differences

Third, we examined whether topic knowledge, academic success in LA, and gender were related to students' interpretations of a persuasive video. The level of topic knowledge seemed, to some extent, to support students' interpretations of the main message of the video. Topic knowledge was associated with the quality of students' reasoning on the usefulness of vaccinations. This is consistent with reading comprehension models of monomodal texts (McNamara & Magliano, 2009) and models of multimodal text processing (Magliano et al., 2013; Verhoeven & Perfetti, 2008) that emphasize the importance of topic knowledge in constructing a coherent representation of a text. Unexpectedly, the higher the student's topic knowledge was, the more likely he or she was to give a negative interpretation of the safety of vaccines. Notably, although students did not select the statement that vaccines cause autism in the topic knowledge test, 19.6% of students interpreted the video as presenting the idea that vaccines may cause autism.

Furthermore, with one exception, topic knowledge did not support students with connecting modes to the main argument(s) of the videos. It might be that other types of prior knowledge, such as cultural knowledge and knowledge about media genres,

play a role when students interpret the use of modes in meaning-making (see Ajayi, 2011). Furthermore, we did not find any consistent patterns between the students' subject grades and their interpretations of the video.

Previous research has found that girls tend to perform better than boys in traditional areas of literacy (Organisation for Economic Co-operation and Development, 2013; Torppa et al., 2018) and digital literacies contexts (Naumann & Sälzer, 2017; Salmerón et al., 2018). In the multimodal interpretation tasks, we found no consistent gender differences. This suggests that there may be slightly different individual and motivational factors that contribute to the reading of multimodal texts. It would be worth examining whether media use, interest in the topic, or medium preferences could moderate girls' and boys' performance in interpreting videos.

Limitations and Future Research

This study has its limitations. First, because students analyzed the video as a part of a larger online inquiry task and we did not want to overwhelm them, we asked them to select and analyze the three modes they considered the most important from a list of five (audio, visual, motion, text, and gestures and facial expressions). Thus, students interpreted somewhat different sets of modes for the task, which limited our possibilities both to calculate overall scores across modes and to examine whether interpreting modes is a latent skill.

Second, in our analyses of students' interpretations of modes, we gave credit when a student identified at least one multimodal element of a specific mode and additional credit when the student connected this element to the main argument(s), topic, or communicative purposes of the video. Thus, our analysis does not reflect the richness of students' descriptions and interpretations of specific modes or how modes were layered and interacted (see Dalton et al., 2015; Towndrow et al., 2013). Furthermore, our analysis considered one mode at a time. In future research, students could be asked to interpret not only how different modes are involved in meaning-making but also the intersemiotic relationship between them (Jewitt, 2009; Kress, 2010).

Third, this study focused on students' interpretations of a multimodal video without considering how students perceived the credibility of the source aiming not only to persuade but also to entertain and how these perceptions of credibility might have influenced their interpretations. We would also like to acknowledge that the study was conducted in a particular educational and cultural context, which poses limits for generalization. For example, in Finland, all schools at primary, secondary, and upper secondary levels follow the same national curriculum. In addition, Finland has a national vaccination program, and according to the Finnish Institute of Health and Welfare, the vaccination coverage of young children is very good.

Finally, our study only covered a few individual difference variables. Although we took into account students' prior topic knowledge, we did not ask about their prior stance on vaccination. Some individuals may have a strong, even emotional stance to vaccination that may influence their interpretations (Greene et al., 2019). We were also

limited to examining the role of students' academic success in their interpretations of multimodal videos through their LA grades. While the video concerned a socioscientific topic, it remains for future studies to examine how science grades would contribute, in particular, to students' reasoning about scientific issues communicated through multiple modes.

Although videos can integrate a variety of modes for meaning-making, the format is also imbued with communicative constraints, and it would be beneficial to learn more about how students interpret a variety of digital multimodal products. It would also be fruitful to learn more about how students interpret different multimodal texts based on their modal density and complexity (Norris, 2009), as well as what cultural knowledge and experiences students draw upon when making their interpretations.

Instructional Implications

Our results suggest that there is a need for explicit instruction in schools that focuses on the interpretation of dynamic, multimodal texts designed to persuade. The analytical framework developed in this study could be applied in crafting cognitive or metacognitive prompts (see van den Boom et al., 2004) aimed at drawing students' attention to different aspects of meaning-making: how multiple modes are used to build an argument, to introduce or illustrate a topic, or for communicative purposes. As others have suggested (Bailey, 2009; Bruce, 2008), it might be beneficial for students tasked with analyzing a complex, multifaceted video to first conduct focused viewings in which they attune to specific modes one at a time. Students can then proceed to analyze the relationships between modes and how they interact to create layered meaning.

As suggested in this study, different students may interpret the meaning of a video differently. Therefore, collaborative learning practices where students can share, discuss, and negotiate their interpretations may be particularly helpful (Teasley, 1995). Creating a multimodal transcript collaboratively (see Online Appendix B) could also be a useful tool for practicing systematic analysis of different modes and help students to recognize how they work together to make meaning. Joint representation would also support students' discussions around the role of specific modes and the relationships between them. Furthermore, students who excel at multimodal analysis could serve as resources in the classroom. The integration of videos into classroom practices may also provide a link between school and informal literacy practices (Erstad, 2012) and may have the potential to engage students who struggle with traditional areas of literacy (Toohey et al., 2012).

Finally, while reading and writing are often intertwined literacy practices, students could benefit from instruction that combines analyzing and composing videos, probably in a reciprocal relationship. Close analysis of modes might give students new ideas that they can apply in producing their own videos and lead them to ponder their design decisions more deeply. On the contrary, students' experiences of multimodal composition could give them insights into the effective use of modes that may be helpful when interpreting videos.

To conclude, providing students with opportunities to interpret and compose multimodal videos could serve as an integral part of media literacy instruction that aims at fostering the knowledge and competencies necessary for full participation in today's media-saturated society (Hobbs, 2019). In her recent work, Hobbs (2019) highlighted core ideas of media literacy, including that all media messages are constructed, different people interpret media messages differently, and media messages are powerful because they can influence people's perceptions, attitudes, and behavior in various aspects of life. We believe that the results of our study suggest that instruction centered on the interpretation of multimodal videos has the potential to provide students with insights into these increasingly relevant issues.

Acknowledgments

The authors are grateful to Timo Salminen and Elina Hämäläinen for their help in data collection and to Maarit Vähä-Piikkiö for her help in data analysis.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the Academy of Finland under Grant Number 285817.

ORCID iD

Carita Kiili https://orcid.org/0000-0001-9189-4094

Supplemental Material

The appendix referenced in this article and abstracts in languages other than English are available at http://journals.sagepub.com/doi/suppl/10.1177/1086296X211009296

References

Ajayi, L. (2011). A multiliteracies pedagogy: Exploring semiotic possibilities of a Disney video in a third grade diverse classroom. *The Urban Review*, *43*, 396–413. https://doi.org/10.1007/s11256-010-0151-0

Alexander, P. A. (2016). The arguments for and the reasoning about epistemic cognition. In J. A. Greene, W. A. Sandoval, & I. Bråten (Eds.), *Handbook of epistemic cognition* (pp. 100–110). Routledge.

Arguel, A., & Jamet, E. (2009). Using video and static pictures to improve learning of procedural contents. *Computers in Human Behavior*, *25*(2), 354–359. https://doi.org/10.1016/j.chb.2008.12.014

Bailey, N. M. (2009). It makes it more real: Teaching new literacies in a secondary English classroom. *English Education*, *41*(3), 207–234. https://www.jstor.org/stable/40173321

Barzilai, S., & Eilam, B. (2018). Learners' epistemic criteria and strategies for evaluating scientific visual representations. *Learning and Instruction*, *58*, 137–147. https://doi.org/10.1016/j.learninstruc.2018.06.002

Blair, J. A. (2015). Probative norms for multimodal visual arguments. *Argumentation*, *29*, 217–233. https://doi.org/10.1007/s10503-014-9333-3

Bruce, D. (2008). Visualizing literacy: Building bridges with media. *Reading & Writing Quarterly*, *24*(3), 264–282. https://doi.org/10.1080/10573560802004126

Cromley, J. G., Snyder-Hogan, L. E., & Luciw-Dubas, U. A. (2010). Cognitive activities in complex science text and diagrams. *Contemporary Educational Psychology*, *35*(1), 59–74. https://doi.org/10.1016/j.cedpsych.2009.10.002

Dalton, B., Robinson, K., Lavvorn, J., Smith, B. E., Alvey, T., Mo, E., Uccelli, P., & Proctor, C. P. (2015). Fifth-grade students' digital retellings and the Common Core: Modal use and design intentionality. *The Elementary School Journal*, *115*(4), 548–569.

Enders, C. K. (2010). *Applied missing data analysis*. Guilford Publications.

Erstad, O. (2012). The learning lives of digital youth—Beyond the formal and informal. *Oxford Review of Education*, *38*(1), 25–43. https://doi.org/10.1080/03054985.2011.577940

Greene, J. A., Cartiff, B. M., Duke, R. F., & Deekens, V. M. (2019). A nation of curators: Educating students to be critical consumers of online information. In P. Kendeou, D. H. Robinson, & M. T. McCrudden (Eds.), *Misinformation and fake news in education* (pp. 187–206). Information Age.

Haria, P. D., & Midgette, E. (2014). A genre-specific reading comprehension strategy to enhance struggling fifth-grade readers' ability to critically analyze argumentative text. *Reading & Writing Quarterly*, *30*(4), 297–327. https://doi.org/10.1080/10573569.2013.818908

Hobbs, R. (2019). Media literacy foundations. In R. Hobbs, & P. Mihailidis (Eds.), *The international encyclopedia of media literacy* (pp. 1–19). John Wiley. https://doi.org/10.1002/9781118978238.ieml0063

Howell, E., Butler, T., & Reinking, D. (2017). Integrating multimodal arguments into high school writing instruction. *Journal of Literacy Research*, *49*(2), 181–209. https://doi.org/10.1177/1086296X17700456

Hull, G. A., & Nelson, M. (2005). Locating the semiotic power of multimodality. *Written Communication*, *22*(2), 224–261. https://doi.org/10.1177/0741088304274170

Hull, G. A., Stornaiuolo, A., & Sahni, U. (2010). Cultural citizenship and cosmopolitan practice: Global youth communicate online. *English Education*, *42*(4), 331–367. https://www.jstor.org/stable/23018017

Iordanou, K., Kendeou, P., & Beker, K. (2016). Argumentative reasoning. In J. A. Greene, W. A. Sandoval, & I. Bråten (Eds.), *Handbook of epistemic cognition* (pp. 39–53). Routledge.

Jewitt, C. (2009). *The Routledge handbook of multimodal analysis*. Routledge.

Kendeou, P., Bohn-Gettler, C., White, M. J., & Van Den Broek, P. (2008). Children's inference generation across different media. *Journal of Research in Reading*, *31*(3), 259–272. https://doi.org/10.1111/j.1467-9817.2008.00370.x

Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge University Press.

Kjeldsen, J. E. (2015). The study of visual and multimodal argumentation. *Argumentation*, *29*, 115–132. https://doi.org/10.1007/s10503-015-9348-4

Kjeldsen, J. E. (2016). Symbolic condensation and thick representation in visual and multimodal communication. *Argumentation and Advocacy*, *52*(4), 265–280. https://doi.org/10.1080/00028533.2016.11821874

Kress, G. R. (2010). *Multimodality: A social semiotic approach to contemporary communication*. Routledge.

Kress, G. R., & Van Leeuwen, T. (1996). *Reading images: The grammar of visual design*. Routledge.

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

Leu, D. J., Kinzer, C. K., Coiro, J. L., & Cammack, D. W. (2004). Toward a theory of new literacies: Emerging from the internet and other information and communication technologies. In R. B. Ruddell, & N. J. Unrau (Eds.), *Theoretical models and processes of reading* (5th ed., pp. 1570–1613). International Reading Association.

List, A. (2018). Strategies for comprehending and integrating texts and videos. *Learning and Instruction*, *57*, 34–46. https://doi.org/10.1016/j.learninstruc.2018.01.008

Magliano, J. P., Loschky, L. C., Clinton, J., & Larson, A. M. (2013). Is reading the same as viewing? An exploration of the similarities and differences between processing text-and visually based narratives. In B. Miller, L. Cutting, & P. McCardle (Eds.), *Unraveling the behavioral, neurobiological, & genetic components of reading comprehension* (pp. 78–90). Brookes.

McNamara, D. S., & Kintsch, W. (1996). Learning from texts: Effects of prior knowledge and text coherence. *Discourse Processes*, *22*(3), 247–288. https://doi.org/10.1080/01638539609544975

McNamara, D. S., & Magliano, J. (2009). Toward a comprehensive model of comprehension. *Psychology of Learning and Motivation*, *51*, 297–384. https://doi.org/10.1016/S0079-7421(09)51009-2

Merkt, M., Weigand, S., Heier, A., & Schwan, S. (2011). Learning with videos vs. learning with print: The role of interactive features. *Learning and Instruction*, *21*(6), 687–704. https://doi.org/10.1016/j.learninstruc.2011.03.004

Mills, K. A. (2010). A review of the "digital turn" in the new literacy studies. *Review of Educational Research*, *80*(2), 246–271. https://doi.org/10.3102/0034654310364401

Muthén, L. K., & Muthén, B. O. (2017). *Mplus user's guide* (8th ed.).

Naumann, J., & Sälzer, C. (2017). Digital reading proficiency in German 15-year-olds: Evidence from PISA 2012. *Zeitschrift für Erziehungswissenschaft*, 1–19. https://doi.org/10.1007/s11618-017-0758-y

Newell, G. E., Beach, R., Smith, J., & VanDerHeide, J. (2011). Teaching and learning argumentative reading and writing: A review of research. *Reading Research Quarterly*, *46*(3), 273–304. https://doi.org/10.1598/RRQ.46.3.4

New London Group. (1996). A pedagogy of multiliteracies: Designing social futures. *Harvard Educational Review*, *66*(1), 60–93. https://doi.org/10.17763/haer.66.1.17370n67v22j160u

Norris, S. (2009). Modal density and modal configurations: Multimodal actions. In C. Jewitt (Ed.), *The Routledge handbook of multimodal analysis* (pp. 78–90). Routledge.

Organisation for Economic Co-operation and Development. (2013). *PISA 2012 results: Excellence through equity (Volume II): Giving every student the chance to succeed*. https://doi.org/10.1787/9789264201132-en

Paivio, A. (1990). *Mental representations: A dual coding approach*. Oxford University Press.

Raykov, T., Dimitrov, D. M., & Asparouhov, T. (2010). Evaluation of scale reliability with binary measures using latent variable modeling. *Structural Equation Modeling*, *17*(2), 265–279. https://doi.org/10.1080/10705511003659417

Reznitskaya, A., Anderson, R. C., & Kuo, L. J. (2007). Teaching and learning argumentation. *The Elementary School Journal*, *107*(5), 449–472.

Rideout, V. J., Foehr, U. G., & Roberts, D. F. (2010). *Generation M2: Media in the lives of 8- to 18-year-olds*. Kaiser Family Foundation. https://www.kff.org/wp-content/uploads/2013/01/8010.pdf

Salmerón, L., García, A., & Vidal-Abarca, E. (2018). The development of adolescents' comprehension-based Internet reading skills. *Learning and Individual Differences*, *61*, 31–39. https://doi.org/10.1016/j.lindif.2017.11.006

Salmerón, L., Sampietro, A., & Delgado, P. (2020). Using Internet videos to learn about controversies: Evaluation and integration of multiple and multimodal documents by primary school students. *Computers & Education*, *148*, Article 103796. https://doi.org/10.1016/j.compedu.2019.103796

Serafini, F. (2010). Reading multimodal texts: Perceptual, structural and ideological perspectives. *Children's Literature in Education*, *41*, 85–104. https://doi.org/10.1007/s10583-010-9100-5

Shanahan, L. E. (2013). Composing "kid-friendly" multimodal text: When conversations, instruction, and signs come together. *Written Communication*, *30*(2), 194–227. https://doi.org/10.1177/0741088313480328

Smith, A., & Anderson, M. (2018, March 1). *Social media use in 2018*. Pew Research Center. http://www.pewinternet.org/2018/03/01/social-media-use-in-2018

Smith, B. E., Kiili, C., & Kauppinen, M. (2016). Transmediating argumentation: Students composing across written essays and digital videos. *Computers & Education*, *102*, 138–151. https://doi.org/10.1016/j.compedu.2016.08.003

Smith, B. E., Pacheco, M. B., & Khorosheva, M. (2021). Emergent bilingual students and digital multimodal composition: A systematic review of research in secondary classrooms. *Reading Research Quarterly*, *56*(1), 33–52. https://doi.org/10.1002/rrq.298

Stein, P. (2008). Multimodal instructional practices. In J. Coiro, M. Knobel, C. Lankshear, & D. J. Leu (Eds.), *Handbook of research on new literacies* (pp. 871–898). Lawrence Erlbaum.

Strauss, A., & Corbin, J. M. (1998). *Basics of qualitative research: Techniques and procedures for developing grounded theory* (2nd ed.). SAGE.

Teasley, S. D. (1995). The role of talk in children's peer collaborations. *Developmental Psychology*, *31*(2), 207–220. https://doi.org/10.1037/0012-1649.31.2.207

Toohey, K., Dagenais, D., & Schulze, E. (2012). Second language learners making video in three contexts. *Language and Literacy*, *14*(2), 75–96. https://doi.org/10.20360/G2S59R

Torppa, M., Eklund, K., Sulkunen, S., Niemi, P., & Ahonen, T. (2018). Why do boys and girls perform differently on PISA reading in Finland? The effects of reading fluency, achievement behaviour, leisure reading and homework activity. *Journal of Research in Reading*, *41*(1), 122–139. https://doi.org/10.1111/1467-9817.12103

Toulmin, S. (1958). *The uses of argument*. Cambridge University Press.

Towndrow, P. A., Nelson, M. E., & Yusuf, W. F. B. M. (2013). Squaring literacy assessment with multimodal design: An analytic case for semiotic awareness. *Journal of Literacy Research*, *45*(4), 327–355. https://doi.org/10.1177/1086296X13504155

Tseronis, A. (2018). Multimodal argumentation: Beyond the verbal/visual divide. *Semiotica*, *2018*(220), 41–67. https://doi.org/10.1515/sem-2015-0144

UltraMiraculous. (2010, August 20). Penn and Teller on Vaccinations [Video]. *YouTube*. https://www.youtube.com/watch?v=RfdZTZQvuCo

Unsworth, L. (2008). Multiliteracies and metalanguage: Describing image/text relations as a resource for negotiating multimodal texts. In J. Coiro, M. Knobel, C. Lankshear, & D. J. Leu (Eds.), *Handbook of research on new literacies* (pp. 377–405). Lawrence Erlbaum.

Van den Boom, G., Paas, F., Van Merrienboer, J. J. G., & Van Gog, T. (2004). Reflection prompts and tutor feedback in a web-based learning environment: Effects on students' self-regulated learning competence. *Computers in Human Behavior*, *20*(4), 551–567. https://doi.org/10.1016/j.chb.2003.10.001

Verhoeven, L., & Perfetti, C. (2008). Advances in text comprehension: Model, process and development. *Applied Cognitive Psychology*, *22*(3), 293–301. https://doi.org/10.1002/acp.1417

Walsh, M. (2006). The "textual shift": Examining the reading process with print, visual and multimodal texts. *Australian Journal of Language and Literacy*, *29*(1), 24–37.

YouTube. (n.d.). *YouTube by the numbers*. https://www.youtube.com/yt/about/press/

Manuscript submitted: 26 February 2020
Manuscript accepted: 15 September 2020
Article published online: XX