

MULTIMODAL ANALYSIS OF MATHEMATICS VIDEO TUTORIALS: ORCHESTRATION OF MODES

Brno Studies in English
Volume 48, No. 2, 2022

ISSN 0524-6881 | e-ISSN 1805-0867
<https://doi.org/10.5817/BSE2022-2-2>

SIMONA KORYTÁŘOVÁ

Abstract

Based on the multimodal discourse analysis, the contribution attempts to describe the genre of online video tutorial, focusing primarily on the involved modes and their interplay. The research has been done on the corpus consisting of online video tutorials on differential and integral calculus (e.g. first-order linear differential equations) which are available on the Internet and made by native speakers of English. In order to comment on differences between online and offline practices a referential group of standard lectures on infinitesimal calculus has been investigated as well. A case study focusing on forms of explication from the multimodal perspective has been carried out; seven presentational formats were compared both quantitatively and qualitatively to find out how various communication techniques are employed.

Key words

Educational videos; multimodality; multimodal ensemble; instructional design; video tutorial

1. Introduction

In recent years, remote tutoring has spread thanks to the availability of varied means of computer technology; there is a growing number of educational videos to be found on video-sharing websites, they are also employed in a broad range of digital learning environments such as Khan Academy and MOOCs. Tutorials can take many forms; it is a genre without strict rules, so their producers can use various techniques to convey the message. This broad variety of design strategies makes online tutorial an interesting subject of linguistic research. Not only do video tutorials vary in respects such as length, quality, and used equipment, they also show a relatively wide range of modes employed. A detailed investigation of the way producers of tutorials make use of modes might help understand how the communication via platforms such as YouTube is carried out and made effective.

Digital technology usage has recently attracted a great deal of attention in various fields of research; concerning linguistics, Jewitt (2016) notes that digital technologies are of particular interest to multimodality because the wide range of modes which they make available often appear in new inter semiotic relationships with one another. She claims that “they unsettle and re-make genres, and they

often reshape practices and interaction” (2016: 70). This unsettling can be illustrated by showing the shift of spatial and temporal boundaries of communication while using digital technologies. A face-to-face conversation in which participants share time and space changes into a video chat where only time is shared and participants are physically separate and this transformation can proceed further to an online video tutorial where the communication is detached both spatially and temporally but still an imaginary recipient is implicitly present in the interaction and thus influences its structure.

Digital technologies allow a wide variety of techniques to be used; since more modes can be employed it brings about their increased usage in educational environment. Accordingly, there has been a long-time interest in the relation between multimodality and education (Kress, 1998, 2003; Kress et al., 2001; O’Halloran, 2005, 2015). Bezemer and Kress (2015) stress the importance of taking all modes of representation seriously in the school curriculum; staying away from privileging language over other modes is highly recommended also by Norris (2004). The meaning-making should be understood as a whole; this approach, i.e. moving away from a sole focus on language, can be seen in many recent studies, some of which will be mentioned in the following paragraphs.

Several studies concerning instructional design have been focused on the modality effect which relates to the cognitive capacity being increased by the simultaneous use of auditory and visual channels: e.g., van den Broek et al. (2014) investigated how learner performance depends on the presentation mode of studied items and how the modality effects change over time. Kruger and Doherty (2016) propose a multimodal methodology to provide a framework for the measurement of cognitive load in the presence of educational video; in this study, they are particularly interested in verbal redundancy and its impact on managing cognitive load.

Other specific features of the multimedia design have been the focus of many studies as well. Considerable attention has been paid to instructor presence (Wang & Antonenko, 2017) and more specifically to instructor presence features such as dynamic drawings, eye contact with the camera, and instructor visibility (Fiorella et al., 2018). Kizilcec et al. (2014) also report on the value of seeing the lecturer’s face in a lecture design, while Beege et al. (2017) investigate the effect of addressing in educational videos by manipulating orientation and proximity. Stull et al. (2018) discuss use of transparent glass boards which allows the instructor to face the students while writing and providing a concurrent explanation of the material. They suggest it promotes learning as the affordances of transparent whiteboard lessons might better follow design principles of multimedia learning and foster social agency compared to conventional whiteboard lessons.

As far as investigations of science classroom discourse are concerned, some of them have dealt with topics of multimodal meaning making (Danielsson, 2016) and instantiation of specific scientific meaning (Tang, 2013); Tang et al. (2014) analyse student representation practices using their integrative framework for the analysis of multiple and multimodal representations. O’Halloran (2015: 73) focuses on the multimodal register for mathematics and points out the importance of multimodal literacy described as “a literacy which extends beyond lan-

guage to include mathematical symbolic notation and mathematical images". The multisemiotic discourse of mathematics is thoroughly discussed in *Mathematical Discourse* (2005) where O'Halloran examines the grammars of mathematical symbolism and mathematical visual images, and explores intersemiosis between language, visual images and symbolism.

In recent years, the genre of education videos has attracted attention of several researchers. Bateman and Schmidt-Borcherding (2018) contribute to the cognitive research in multimodal learning by focusing on communicative effectiveness of education videos and interweaving empirical investigation and theories of multimodal discourse. They compare three education videos selected to represent the three basic video lecture presentation styles defined by Chorianopoulos and Giannakos (2013) and investigate empirically how their presentation and development of discourse entities are managed via the presence or non-presence of audio-visual macro-themes.

Various presentation styles are the focus of a study carried on by Crook and Schofield (2017). The concept of 'lecture' is concerned here as realized in both offline and online contexts; Crook and Schofield also explore how media differences entail different student experiences. They introduce a taxonomy of online lecture design, identify five categories and illustrate them by discussing sixteen formats; for example, one of them is termed *Khan whiteboard* as Sal Kahn's name is associated with video design where a voice is narrated over a problem-solving illustration.

O'Halloran (2015) stresses the importance of multimodal approach as it has implications for teaching as well as student learning; she points out that multimodal literacy concerns the relations between three sources, i.e. language, symbolism and images, and their integration. Bezemer et al. (2012: 12) claim that "explication is not the exclusive domain of the modes of speech and writing"; they state that "in many contexts, image, or gesture, are modes better suited to 'make explicit', or even the only modes available to make anything explicit at all". Morell (2018) suggests that effective pedagogy has much to do with multimodal competence, i.e. the ability to understand the combined potential of various modes for making and eliciting meaning. Obviously, the ability to combine modes effectively is useful not only in pedagogy but also for meaning-making in every interaction, in all communication.

Not surprisingly, the opinion that communication is multimodal is held by many scholars. Bezemer et al. (2012) state that in all communication meanings are made in modal ensembles, which consist of various modes orchestrated in complex conjunctions. Norris also claims that all interactions are multimodal (2004); however, what can be classified as a mode is not clear as she notes that the concept of mode is problematical, and modes can be delineated in various ways (2013). According to Jewitt (2016), modes are not autonomous and fixed, but being created through social processes they are fluid and subject to change. Additionally, modes are not universal but are specific to a community where understanding of their semiotic characteristics is shared. This feature is also stressed by Kress, who defines mode as a socially shaped and culturally given resource for making meaning (in Jewitt 2009).

Besides other directions of investigation, multimodal research focuses on the interplay among modes, it views the modes as individual entities and examines their interaction with other modes. There is a broad range of modes that have been identified, e.g. speech, writing, gaze, gestures, posture, images, colour. After years of being sidelined by mainstream linguistics, non-verbal modes have recently been investigated more frequently and attention has been paid to a wider repertoire of communicational means. The variety of ways in which meanings are made is explored; identifying new means of interaction offers further possibilities of more detailed analyses and finding out how the meanings are multimodally interrelated. Ways of approaching this matter are numerous; e.g. even the presence of face can be considered an important and thus researchable property of interaction. This research was done by Kizilcec (2015), who points out that the primary mode of knowledge transfer has been face-to-face communication; the familiarity of seeing a face during instruction influences positively the process of paying attention as it was found that hearing the speech and seeing the instructor's face alleviate learners' cognitive load.

Appearing in numerous variations and thus being quite complex in the terms of their modal diversity yet remaining simpler than more technologically advanced communicative media such as virtual and augmented reality, educational videos seem to be convenient to be investigated from the multimodal point of view. This approach is also recommended by Bateman and Schmidt-Borcherding (2018: 2) who state that “educational videos constitute a highly beneficial test case for the development of multimodally more sophisticated analytic frameworks and their related empirical methods”. Educational video lectures from many disciplines are made broadly accessible via platforms such as YouTube; their forms vary considerably since each format deals with the explication of a task in its own particular way. As Bateman and Schmidt-Borcherding (2018) point out, online educational video constitutes a complex form of multimodal communication; to illustrate its complexity the range of modes employed in Sample 1E6 will be briefly described in the following paragraphs.

Sample 1E6 is a nine-minute-long video tutorial called *Chain rule for derivatives, with product rule*. The video tutorial consists of three parts: introduction, body, and ending. The introduction is quite short, it lasts only 20 seconds; it features the tutor welcoming users and informing them about the subject matter of the tutorial. The tutor is visually present on the left side of the screen; as only her head and shoulders are visible, there are no gestures involved. However, she employs gaze and mimics to establish contact with receivers; on the right side there are several words appearing gradually as they are pronounced: chain rule, identify, derivative, multiply. They name the method and mark the three stages of the task she is going to demonstrate. Since it is the spoken language that develops the content, the written single words could be argued to be informationally redundant, but they are certainly not redundant with respect to their didactic role.

The major part of the tutorial is occupied by showing the application of the chain rule to this mathematics task: Use Chain Rule to find the derivative of $g(x) = (1+4x)^5 (3+x-x^2)^8$. The tutor explains thoroughly how to use the chain rule to calculate the derivative of the product of two functions. As the tutor is not visible

in the main part of the video, she can use only her voice and employ the properties of this “tool”, such as intonation and loudness. The explication is gradually written on a digital blackboard with a writing tool in the shape of a chalk which produces a white text. The tutor starts writing in the upper-left corner of the screen and follows the top-down and left-right direction. In the end, the blackboard is filled with a short text, which is in the form of mathematical notation. As the whole task is solved in seven lines, it can be presented in one screen.

Short parts of the text are underlined by four different colours in order to distinguish between two functions which are multiplied and later to mark the inside functions while explaining the method. Despite the brevity of the text, several forms of written representations can be identified: there are only several words, most of the text is formed by mathematical notation consisting of numbers, signs, and letters representing variables and functions. Regarding the written words, only 18 of them are used in the whole video: setting the task, key words, and link. The previously mentioned forms can be considered first-order visual resources; second-order visual resources used for navigation, such as underlining and crossing out, serve as didactic means which facilitate understanding. Another mode which can be identified in this sample is layout; the layout of the mathematical notation is clearly arranged, and it adheres to the standard followed by the community, which helps understanding. Since the tutor is not visible, there is no explicit attention guidance with bodily gestures, such as pointing, towards the visual presentation. Instead, the chalk-shaped writing tool is used to point at the text.

Further means of communication employed in this sample are hyperlinks. After clicking on the letter “i” in the upper-right corner of the screen, a small window presents a small photo of the tutor and a link to her website, below which there is a text saying “Step-by-step Derivates course” followed by another line saying “learn more” which can be clicked on. After clicking on the last line, a website is open; this time there is a large photo of the tutor sitting on a chair next to a table, which is placed in the centre of the photo. In the middle of the screen there is a question “Want to master derivatives? I have a step-by-step course for that.” This text is written in four lines and placed in the centre of the screen as it represents the key message. Below the question and offer, a distinct red oval “button” with two words written in capital letters says “GET STARTED”. Unlike the small photograph, in this photo the tutor looks directly at the viewer. It might be a gradual invitation: the first step features the tutor looking away with a smile, the second step shows her keeping the smile and adding a gaze working as a direct offer of contact. The homepage can be scrolled down, and it provides information concerning the courses offered. Another attempt to redirect users to this website is made after one minute of the video when a notification Get help with Derivatives is shown in the upper-right corner for several seconds; if clicked on, it provides you again with the information described above.

The ending is very short, it consists of eight-second-long speaking and seven seconds of showing a link. The tutor appears again and concludes by saying “So I hope you’ve found that video helpful if you did like this video down below and subscribe to be notified of future videos”. While saying that, two words appear

one after another written to the right of her head: first *like* and then *subscribe*. The second one is placed higher and is significantly bigger than the other one which might suggest what step is favoured by the tutor. The tutorial is ended by showing a link to the tutor's website.

To sum up, the video combines two formats categorized by Crook and Schofield (2017): introduction and conclusion are in the *presence-in-picture* format, while the body of the video is made in *Kahn-whiteboard* format. The modes employed here could be identified as speech, writing (including underlining), colours, layout, pointing, gaze. Furthermore, there is also an intertextual component – hyperlink, which refers to another source of information.

Taking into consideration the range of modes that can be employed in multimodal ensembles, the ways of combining modes are numerous; to find the efficient compositions which support explication and facilitate comprehension is the task tutorial producers are confronted with. Bateman and Schmidt-Borcherding (2018) observe that there is surprisingly little empirical work on the instructional design of education videos, and they point out that it needs to be explored which multimodal structures are appropriate to support learning with educational videos. In order to attend to this exploration of educational landscape, a case study focusing on forms of explication from the multimodal perspective has been carried out.

The first and obvious step in the multimodal analysis of the investigated set is identification and classification of the modes involved. After being classified the modes should be assessed in relation to their function both quantitatively and qualitatively; a primary or dominant mode should be determined since it can reveal the structure of video tutorial. The characterisation of relations between modes, e.g. to what extent they are used simultaneously, might help to classify various presentational styles and to explain how the analysed modes cooperate with the others in the multimodal ensemble.

2. Material

For the purpose of this case study a set of samples was taken from the corpus of lecture recordings and video tutorials available on YouTube. Both lectures and tutorials deal with topics from infinitesimal calculus, such as implicit differentiation, integration techniques, methods of solving differential equations etc. The structure of the original corpus and a detailed description of the case-study samples are presented in the following subchapters.

2.1 The corpus

The corpus includes four basic sets of samples; each set deals with a different topic from differential and integral calculus. In Set 1 basic techniques of differentiation – chain rule and product rule – are taught. Set 2 deals with implicit differentiation, which can be considered as a higher level of differentiation. Set 3, differential equations, contains tutorials on a comparatively difficult topic, which

requires involving integrals and mastering of several strategies. For this set tutorials on the usage of the integrating factor at solving first order linear differential equations were chosen. In Set 4 u-substitution, a basic integration technique, is being explained. When the content is taken into consideration, the topics are similar enough to ensure the homogeneity of data. On the other hand, the fact that two of them are simple and thus easily explainable and the others require more effort to be put into the explanation should guarantee that a wider range of speaking practice is covered.

The corpus consists of two subcorpora; the first one comprises four lecture excerpts, their total length is 40 minutes. The subcorpus of online video tutorials includes 41 tutorials, their total time is 5 hours 30 minutes. There are 21 speakers in the corpus: 16 men and 5 women.

The samples are divided into five groups (marked A, B, C, D, and E) in accordance with two criteria. The first one was used to make the two subcorpora: in the first collection (i.e. in Group A) the audience is present, the second set (i.e. Groups B, C, D, and E) is characterized by the absent audience. The second criterion is based on the degree of similarity to a standard lecture; it means that the conditions in B are more similar to A than conditions in C compared to A. In other words, the setting of tutorials in Group B resembles a lecture the most; Group C can be characterised by an unusual setting, for example two windows, one with an emerging text and the other showing only the speaker's head, or the speaker is standing behind a glass board and keeping eye contact while writing on the glass; in Group D only the speaker's hands are visible, while in Group E there is no part of the speaker's body visible.

2.2 The case study

The case study consists of eight samples: one is taken from the subcorpus of lecture recordings, seven from the subcorpus of online video tutorials. There are six tutors, four men and two women, both professional and amateur mathematics teachers. In this set only one topic is being explained – differential equations. The total time of eight samples is 60 minutes, spoken text consists of approximately 9,200 words. The samples were chosen to represent all the five groups; they are characterised briefly in the following paragraphs.

Sample 1 is taken from Group A, which collects recordings of lectures, and it is used in this case study as a referential sample. The lecturer obviously knows that his lecture is recorded as it is an official video posted by MIT; however, a middle part of the lecture was chosen to be analysed, therefore at that point the speaker could have got used to the fact his performance is recorded and he behaves naturally. Sample 1 represents face-to-face communication, so it is the sample which features all the modes identified in this case study.

Sample 2 belongs to Group B, which is the first group of video tutorials. A teaching assistant at MIT is giving instructions how to solve a task; the main difference from Sample 1 is the absence of audience. As the tutor is speaking to an imaginary receiver, it results in a slightly different posture and leads to her looking longer to the camera.

Sample 3 is also taken from Group B; however, it differs from the previous sample mainly in the setting. The tutor is writing on a whiteboard, but she is sitting next to it, so she has much less room to move. She tries to keep the contact with receivers by looking at the camera but as she is from the side, she must turn her head. She is able to combine several modes at the same time, for example she often speaks and points at the text and looks at the imaginary recipient.

Sample 4 features the same tutor, Krista, but it belongs to a different group, i.e. Group C, because the setting differs much more from the lecture setting, which is a default setting for the purpose of this case study. The written text is prepared in advance, it is gradually revealed, always by showing the whole line of the task. The tutor can be seen in a small window; unlike Sample 3, it is a front view. In the previous sample, Krista pointed by using her finger, here she uses a cursor in the shape of circle, and she employs a circular movement of the cursor more often. Almost one third of the time, she looks at the recipient, which is not much different from an ordinary conversation. She also combines three modes as she does in Sample 3, but it is only combination of speaking, gaze and hand gesture, she does not combine speaking and gaze with pointing in Sample 4.

Sample 5 belongs to Group C, so it is similar to the previous one: there is a task which is being solved on the screen and it is accompanied by a small window with the tutor. However, there are several differences. First, it is not front view; second, the tutor makes the text by writing it by hand; third, he points either by a finger or by the pen and sometimes by both simultaneously. He uses gaze quite a lot, but he combines it only with speaking, hardly ever with pointing and never with hand gestures. Hand gestures are rare in this sample even though the conditions are very similar to those from the previous sample, in which the other tutor uses hand gestures quite often.

Sample 6 belongs to Group D, the only part of the tutor's body which can be seen are his hands. The paper he writes on is seen from his viewpoint, so receivers might feel more involved into the process of solving the task. The tutor mostly points by a pen, hand gestures are not frequent.

Sample 7 belongs to Group E whose main feature is visual absence of the tutor. This sample was produced by Krista, who made also Sample 3 and Sample 4; here she writes on a digital blackboard with a cursor in the shape of a chalk. She also uses the cursor for pointing and making circular movement.

Sample 8 is from Group E as well, so it resembles the previous setting; nevertheless, there is one important difference: there is no cursor that could be used for pointing. The tutor tries to compensate by underlining instead of pointing, but this technique cannot be used much as it would make the text confusing.

2.3 The modes observed in the samples

The range of modes used in the investigated samples varies significantly; moreover, the change they exhibit is gradual as shown in Table 1. In Sample 1, all the modes classified in this table are present (i.e. it works here as a default setting) and then a gradual disappearance of the modes can be observed. Besides modes which are commonly identified by multimodal analysis, another mode, a mean-

ingful silence, was added: it is used by the lecturer to highlight what he has said. The reasons for considering meaningful silence a mode are given in 2.4.

This particular set of tutorials on differential equations was chosen from the corpus because it forms a scale and therefore it has a potential to reveal any relations between conditions under which the samples were produced and the modes or the combinations of modes. Another specific feature of this set is the subset of three tutorials made by one tutor which belong to different groups (i.e. samples 3,4, and 7).

Out of the eight modes which have been identified five are chosen to be focused on in this case study: *speaking*, *writing*, *gaze*, *gesture* and *pointing*. The last two were put together in one category called *gesture*, which is then divided into subcategories as it is shown in 4.2.2.

Table 1. The gradual disappearance of modes

	Sample	1	2	3	4	5	6	7	9
	Group	A	B	B	C	C	D	E	E
Mode	speaking	●	●	●	●	●	●	●	●
	writing	●	●	●	●	●	●	●	●
	layout	●	●	●	●	●	●	●	●
	pointing	●	●	●	●	●	●	●	—
	gaze	●	●	●	●	●	—	—	—
	gesture	●	●	●	●	★	—	—	—
	posture	●	●	—	—	—	—	—	—
	silence	●	—	—	—	—	—	—	—

2.4 Is silence a mode?

According to Kress, “mode is a socially shaped and culturally given resource for making meaning” (in Jewitt 2009: 54). Thus, the first question which must be posed is if *silence* makes meaning. As used here by the lecturer, it does: it is different from a pause, i.e. an operational stopping of producing sound in order to take a breath or to concentrate fully on another activity. This kind of silence has its value; the fact there is nothing said is important. This is not a mere pause between two packets of information, it is exactly the opposite since it includes information. Not speaking gives prominence to the previously said by giving time to realize its importance. The meaningful *silence* can be considered a mode as it communicates a certain property of the matter discussed.

Drawing on Hallidayan systemic functional linguistics, Kress provides a test for the status of mode (in Jewitt 2009) which is based on assessing *ideational*, *interpersonal* and *textual metafunctions* in instances of use. The explanation given in the previous paragraph shows that silence used in this way is a semiotic

resource which has the *ideational metafunction*. Regarding the *textual metafunction*, meaningful silence functions aptly in its environment; it is a message entity, coherent both internally and with the environment. It is used either after something important has been said or it can attract attention to something which will be said. As the temporal relation is crucial here, the meaningful silence must be in the closest temporal proximity to the highlighted statement. If it was placed badly, it would not function. Moreover, the connection must be easily understood by receivers, and it needs to be counted as well-acknowledged regularity within the community. If it was changed in its quality, e.g. a sound would be produced instead of silence, it would not work as intended. Consequently, it would not carry out its task and it would not perform the *ideational metafunction* as it would not have the right meaning in the given situation.

Last but not least, the *interpersonal metafunction* must be considered. Can *silence* represent social relations of those engaged in communication? Is it socially shaped? Kress says that “socially, a mode is what a community takes to be a mode and demonstrates that in its practices” (in Jewitt 2009: 58–59). The use of meaningful silence reflects a certain relation between lecturers and students: when the lecturer stops speaking, the students know that it does not indicate the time for turn taking, i.e. they understand their role in the interaction. Furthermore, this understanding is not universal because an outsider might not respond properly. It suggests that this mode is particular to the community which has a shared understanding of its semiotic characteristics.

To sum up, the *ideational metafunction* of the discussed semiotic resource is realized as meaningful *silence* is able to represent what ‘goes on’ in the world, the *interpersonal metafunction* is activated as it constructs social relations and the *textual metafunction* is realized as well since the meaningful silence creates coherence both internally and externally. All the three Hallidayan metafunctions are present, therefore meaningful *silence* can be considered a mode since it passes the *metafunction* test.

3. Method

3.1 Quantitative analysis

For the purpose of this study a simple method has been worked out; it is based on counting words and relating them to the whole, which enables to compare samples of different length. The application of the method is shown in the excerpt below. The text is the spoken text; if the tutor speaks and looks at an imaginary receiver (*speaking + gaze*), the words are highlighted in yellow, *speaking* combined with *writing* is marked in green, *speaking* and *pointing* in pink, other colours mark combinations of three modes, e.g. *speaking + pointing + gaze*. Then the number of words is counted and expressed as a percentage, where 100% is the number of words spoken.

I'm gonna move it to the left side so you'll have **dy over negative three y equals dx** and **it doesn't matter** where the negative three goes because **all that matters**

is that you're pulling y's and x's apart now where anything else in the equation is so as long as we have y on the left and x on the right so let's go ahead now the second step in any differential equation problem is to integrate both sides once we've separated those variables so we've got them separated and now we go ahead and integrate both sides like this so integrating we have here a coefficient it's negative one third right

3.2 Lexical density in mathematical texts

One of the text properties analysed in the present study is lexical density. In general text the lexical density is the number of lexical words divided by the total number of words. As lexical words are considered nouns, adjectives, verbs and adverbs. However, in mathematical texts the content is mediated in a very different way: by numbers, variables, etc. Consequently, the calculation of lexical density in the present study ranks numbers, functions, variables and numerical operations among lexical words.

4. Results

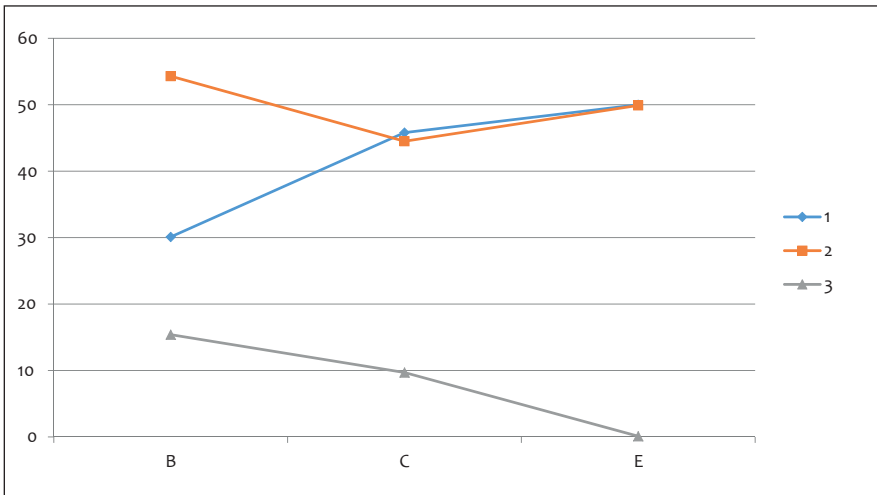
4.1 Concurrent use of modes

In order to describe the employment of modes in various presentational styles, the interplay of modes was to be investigated. The first step of this analysis was assessment of concurrent use of modes. Sample 1 is used in the present study as a referential point located outside the set of tutorials since this sample helped to identify the full set of modes involved. Nevertheless, the modes in Sample 1 were not analysed quantitatively because the lecturer is not visible all the time which means that *gesture* and *gaze* cannot be assessed exactly. The research is focused on the subset of tutorials, i.e. Samples 2–8. The Graphs 1 and 2 compare the investigated modes, it means *speaking*, *writing* (including framing, underlining, and erasing), making *gestures*, and *gaze* in terms of the number of modes used at the same time.

In Graph 1, the data from the Krista subset is compared to determine how the change of setting changes the way of expressing in one person. Three samples from different groups form a scale; when moving from Group B to Group E, fewer modes are available. Line 1 shows changes in employing only one mode, which is *speaking* since the chosen method is based on the number of spoken words. The percentage of a single mode employed is growing up from 30% to 50%, while the parallel employment of three modes is going down from 15% to 0%, which was expected because the concurrent use of *writing* and *gesture* is possible but improbable as it would be rather inconvenient to use the other hand to gesture while writing.

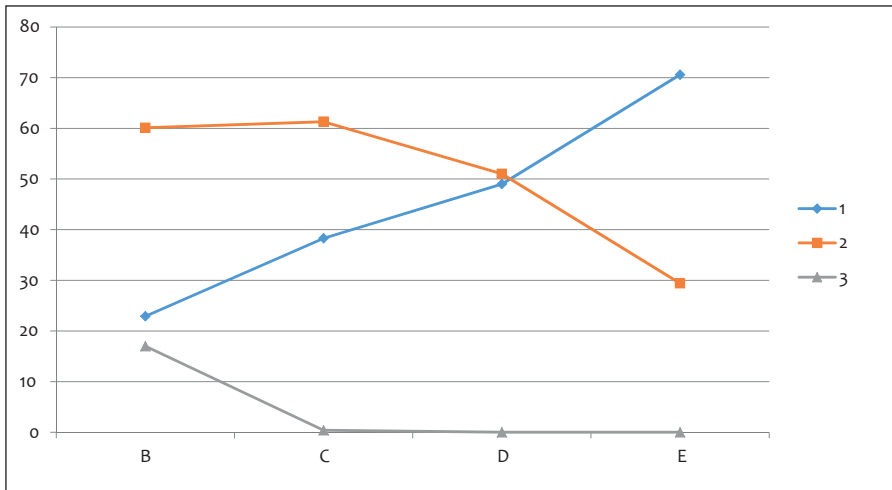
Despite having a wider range of modes in Group B, the tutor prefers using only two of them concurrently – the use of two modes dominates significantly. In Group C, which has the same repertoire of employable modes as Group B but

a different setting, the parallel use of three modes is going down by almost 6%. A decline by circa 10% is observed in the use of two concurrent modes as well, which might be due to the fact the tutor is not used enough to the Group-C technique. Concurrent employment of two modes is almost the same as using only one mode; this relation is also observed in Group E, where the relation between one-mode and two-mode usage is exactly half-and-half.



Graph 1. Concurrent use of modes in the Krista subset

In the subset of the other tutors shown in Graph 2, a similar pattern is followed: firstly, use of two parallel modes dominates over the use of only one mode in Group B and it is significantly higher in Group C as well. However, it must be mentioned that the Group C sample in this subset is rather different from the others as it is produced by the tutor who often (i.e. 23 times during the tutorial) writes without speaking and that is why the use of writing mode does not show in the figures. Nevertheless, he is the only person, who operates this way; all the others do simultaneous *speaking* and *writing*. Secondly, Group D in this graph shows the same relation between one-mode and two-mode usage as Group E in the Krista subset does. When Samples E from these subsets are compared, a significant difference can be seen. Krista employs one mode in 50% of her tutorial, while the male tutor from the other subset shows dominant 70% usage of one mode. However, this is not surprising as he is the one who is most limited in his repertoire of modes because he cannot use a cursor for pointing.



Graph 2. Concurrent use of modes in the other-tutors subset

4.2 Individual modes

4.2.1 Writing

Writing is present in all the investigated samples. There are several properties of this mode which can be discussed; the first quantifiable feature is the process of creating the text. In all the samples, most of the written text is produced gradually: the percentage of prepared text gets over one third just in one sample (see Table 2). This approach might be preferred by all the producers as the gradual building of the text helps reading comprehension by making the text clearer for the receiver through revealing the process of its formation. This feature was also observed and empirically proven by Fiorella and Mayer (2015) whose research suggested that observing the instructor draw diagrams promotes learning.

Another relation, which can be quantified, is a ratio of numbers of the words written and spoken. The resulting fraction is multiplied by 100 to make it easier to imagine (see Table 2). For example, if there are 100 words of written text accompanied by 1000 words of spoken text, the ratio is 10. This comparatively low ratio is observed in the sample from Group A, both samples from Group B and the sample from Group E which was made by Krista. Despite different settings, all samples produced by Krista (i.e. 3,4,7) show low ratio, ranging from 9 to nearly 14; it might mean that her teaching style is primarily based on speaking. Low ratio is also observed in Sample 2, which was produced by the other woman in this study. Samples 5,6, and 8 show twice higher ratio; although their producers are male, the higher number should not be attributed to the gender but to the setting (C, D, E) as the male lecturer from Group A shows the same low ratio as the female tutors.

Finally, the lexical density of the written text can be determined. Taking into consideration the particularities of mathematical text, it is not surprising that the

lexical density of the investigated written texts is extremely high, i.e. on average 96%. This extreme density is “diluted” by the spoken commentary whose lexical density is 46% on average, which decreases the cognitive load of the text and makes the explication more comprehensible.

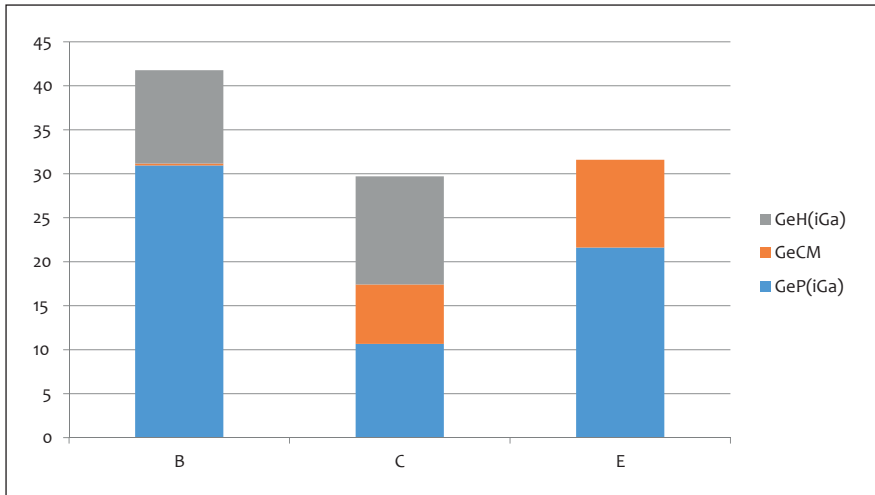
Table 2. Quantification of several features of the written text

Sample	Group	Written text (%)			Ratio W : S × 100	Lexical density (%)
		prepared	gradually appearing			
1	A	0	100		10.96	97
2	B	28.30	71.70		10.73	96.7
3	B	0	100		9.02	100
4	C	0		100	13.73	100
5	C	8.43	91.57		20.82	88.8
6	D	32.17	67.83		21.50	98
7	E	16.84	83.16		10.14	99
8	E	36.36	63.64		21.30	90

4.2.2 Gesture

Writing is an important part of a tutorial since it forms the visual basis of the whole process; the written text is the object which tutors refer to by *pointing*. However, there are more ways of making *gesture* which are used in the investigated tutorials. The three categories identified in the case study are as follows: ordinary pointing (the pointing device is a finger, a hand, a pen, or a cursor), pointing by making a circular movement around one spot, hand gesture other than pointing. It must be noted that in the present study *gesture* means manual gestures, as gesturing with the head is not taken into consideration and analysed.

In Graph 3, the gestures used by Krista are compared. The blue colour represents pointing (GeP), the red is for pointing by making a circular movement (GeCM), and the green colour represents gestures by hands (GeH). Although the use of *gesture* becomes more and more difficult while moving on the scale from lecture-like conditions of Group B to the limited setting of Group E, the observed decrease by 10% is not as distinct as it might be expected. The increase in using the circular-movement pointing is nearly 10%, which might be caused by the producer’s effort to keep the receiver’s attention after losing the visual contact in the Group-E setting. A certain increase can be seen already in the Group-C setting, however in Group-E setting it is 3% higher.



Graph 3. Use of gesture in the Krista subset (Note: iGa means “including gaze”)

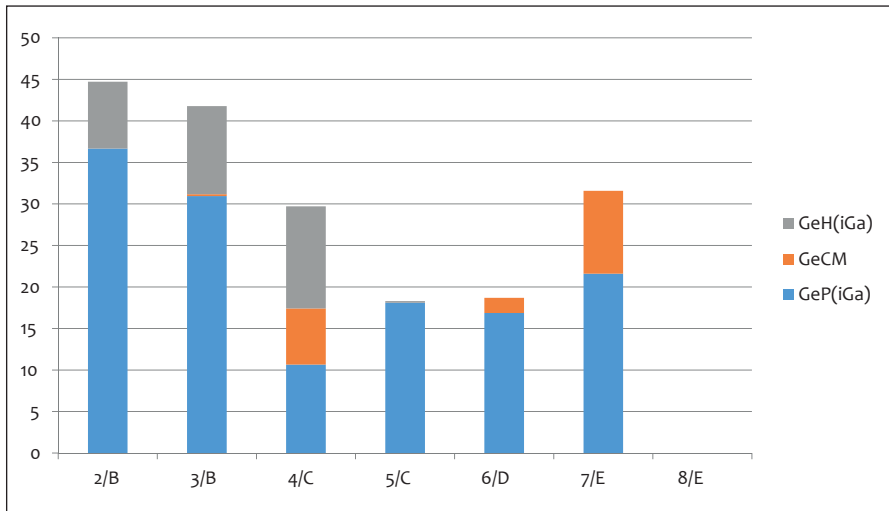
The increasing employment of circular movement deserves a closer look at the function of this tool. Several examples of expressions accompanied by the circular-movement gesture are given below, they are written in italics. Mostly the tutor uses the circular movement to enhance the deixis expressed verbally by words *here, this*; it is shown in examples 1, 2, and 3 where the word in bold is assumed to be the one which is emphasised. It might be used to highlight an expression by a more noticeable device than ordinary pointing.

A rather different occurrence of using this distinct kind of movement is given in example 4, where the tutor makes a circle around an empty space to express the absence of variables. Most likely, it highlights the word *no* while the notion of absence is visualised by drawing attention to the blank space. Thus, the concept of absence is mediated both verbally and visually. The gesture visualises the notion; *speaking* and *gesture* cooperate, moreover, their concurrent use is not informationally redundant as it enhances the information given.

- (1) we have *y prime the derivative of y out **here** by itself*
- (2) if you have *a negative sign **here** in front of your p of x function*
- (3) technically what this means is the derivative of one over x y *but performing **this operation** pulling the d over dx out in front of **this term** here actually takes care of the entire left-hand side and now we just set this equal to sine of x and I'll show you why this works*
- (4) there's **no** y's or *y primes* left on the *right-hand* side

In Graph 4, use of gestures in all the tutors is shown. It can be seen that using hands to make gestures forms a significant part of explaining how to solve a task

in the first three samples; it ranges approximately between 30% and 45%. In Sample 7, hands are replaced by a cursor; however, the tutor is able to use this device both frequently and effectively.



Graph 4. Use of gestures in all the tutors

4.2.3 Summary of modes

Graph 5 sums up the changes in using the investigated modes. The whole column is always based on *speaking*, so the speaking mode either alone or combined with the other modes makes 100%. *Writing* reaches at most 30% (see Sample 8), while *speaking* employed as a single mode is gradually gaining prominence by going from 23% in Sample 2 to 70% in Sample 8.

The total occurrence of *gesture*, which comprises *gesture* and *gesture* combined with *gaze*, ranges between 18% and 45%. Group-B samples show rather a high level of *gesture* usage approaching half of the overall production: in Sample 2 it reaches 45%; similarly, Sample 3 gets only slightly lower to 42%. If the two-modal and three-modal variants are compared in the samples discussed, the difference is not substantial as well: in concurrent *speaking* and *gesture* it is 2% and in *speaking*, *gesture*, and *gaze* only 1%. The higher incidence of *gesture* might be due to the lecture-like setting of both samples which enables making movements which are common in the off-line environment. In this respect, the samples are similar despite rather different conditions of their settings, namely the tutor’s position in relation to the board and their distance from the recording device.

By contrast, Group C is not as homogeneous as Group B: Sample 4 shows 30% of *gesture* usage while Sample 5 only 18%. The observed decline in both samples might be caused by the increased employment of digital technologies which places demands on the speakers while handling the production of educational videos. The considerable drop of *gesture* usage in Sample 5 in comparison with

the other sample from the same group could be attributed partly to a different position of the tutor in relation to the recording device and partly to his teaching style.

The incidence of *gesture* in Sample 6, which represents Group D, is not much different from Group-C Sample 5 as it gets 19%. However, this one is solely a two-modal variant, since its setting does not enable *gaze*. Paradoxically, not having this mode at his disposal might have helped the producer to focus on the explication itself and not to expend his operational capacity on keeping visual contact. Moreover, it seems that the tutor can employ *gesture* more effortlessly because the setting is more natural in comparison with Sample 5 as the camera is placed above the speaker and the setting resembles the off-line format of giving explication to a receiver sitting next to the tutor.

Considering the gradual decrease of the *gesture* employment in samples belonging to groups B, C, and D, a lower value in Group-E samples might be expected; nevertheless, Sample 7 shows rather high value of 32%. This sample was produced by Krista, who made also Sample 3 and Sample 4; despite having different settings they do not evince significantly different *gesture* usage as it ranges between 30 and 42%. Unlike the high figure relating to Sample 7, the zero *gesture* use in Sample 8 is expectable because there are no tools for making gestures available in this setting.

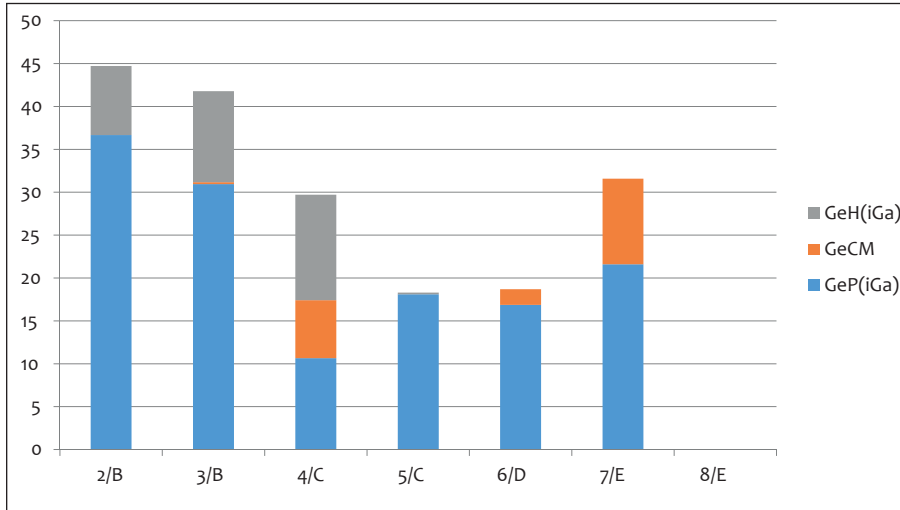
Total *gaze*, which means *gaze* and *gaze* combined with *gesture*, ranges between 23% and 40% in groups B and C; groups D and E do not feature this mode. The observed difference in usage might be influenced mainly by different settings and the resultant posture of tutors, which can be illustrated by comparing the two Group-B samples. In Sample 2 *gaze* reaches 40%; since the tutor is standing in front of the blackboard, she is able to change her position easily. However, in Sample 3 the tutor is sitting in front of the whiteboard, which makes her movement not so flexible; her moves need more effort to be carried out and that might be the reason for the lower occurrence of *gaze*, which gets to 23%.

On the other hand, the two representatives of Group C, Samples 4 and 5, show similar figures (31% and 29% respectively) although the setting is not the same as the tutors face the video camera at different angles. Consequently, the tutor in Sample 4 holds a more convenient position while the one in Sample 5 must turn his head in order to face the recording device. The tutor in Sample 4 uses *gaze* concurrently with *speaking* in 21% of the recording. Since her face is oriented directly to the video camera and the screen, she can follow the emerging text, therefore she is also able to look at the receivers while using *gesture*, which results in additional 10% use of *gaze* and *speaking* combined with *gesture*.

By contrast, the tutor's position in Sample 5 does not allow easy simultaneous looking at the screen and at the camera, so he combines only *gaze* and *speaking*. Nevertheless, he employs it quite often as its occurrence reaches 29%. The concurrent use of *gaze*, *speaking* and *gesture* is rare here, i.e. less than 0.4%, which has been expectable considering the setting used.

Finally, the Group-B samples show the same level of concurrent use of *gaze* and *gesture*, reaching 17% in Sample 2 and 16% in Sample 3; its level in Group C goes down to 10% in Sample 4 and further in Sample 5 it drops to a mere 0.4%.

Even though the difference in simultaneous employment of these modes is not considerable, it suggests that the application of digital technologies makes the simultaneous use of *gaze* and *gesture* rather inconvenient.



Graph 5. Summary of modes (Note: In Samples 3 and 6, erasing of the text is missing, so the columns do not reach 100%.)

5. Conclusion

The present study aims to contribute to the growing body of research on multimodality, higher education and instructional design by better understanding of the differences between traditional and online lecturing. The presented case study focused on multimodal complexity of educational videos involving spoken language, written language and non-verbal communication of various kinds, such as gestures, pointing etc. In order to understand how such communicative forms function, it has been necessary to identify the modes employed, to characterize them and to look into their interplay. The concurrent use of modes has been investigated to find patterns of usage and to determine which mode is predominant.

First of all, it was observed that *speaking* is the mode which is constantly present in the investigated samples. There are only two exceptions: one in the subset of tutorials, where the tutor frequently writes without speaking, and the other one can be found in the lecture where the not-speaking is used to stress the importance of the previously said by silence. All the other speakers never stop speaking, which might have been caused by their effort to avoid “dead air”. The constant producing of sound seems to be a means of holding the receiver’s attention and maintaining the relation between the participants of the act.

Even though it is the spoken language that develops the content, the other modes employed are not informationally redundant. Especially written text is

indispensable in the analysed educational videos because it is used as a skeleton of the whole product and the explanation could not work without it. Since the two modes present in all the investigated samples are *speaking* and *writing*, they can be considered the key modes. However, it seems that *speaking* is the dominant mode as it is the mode which gives the added value to the plain written text which would obviously function on its own, but it is the commentary which is sought for by the users. The predominance of *speaking* is confirmed also quantitatively by the increasing employment of speaking as a single mode rising from 23% to 70%. By contrast, the mode of *writing* does not capitalize on gradual disappearance of the other modes to the extent as the single-mode *speaking*; the maximum value of *writing* usage differs from the arithmetic mean twice less than it does in the case of speaking (i.e. 12.4 and 26.8 respectively).

The preference for gradual building of text which was noted in all the investigated tutors is understandable because it facilitates processing the explication by receivers. The observed tendency to more frequent concurrent use of only two modes despite feasibility of three-mode usage in groups B and C might be attributed to the effort to prevent perception overload of receivers. Interestingly, two most popular and experienced tutors from this case study, Krista and Patrick, adhere to the same pattern in Sample 6 and Sample 7. Although the samples belong to different groups, they both show a 50/50 proportion of one-mode and two-mode use; it suggests that this balanced ratio might be the effective “dosage” of concurrent use of modes which could provide a convenient learning environment.

One of the tutors, Krista, provides a rare opportunity to compare three instructional styles as her tutorials form a subset of three videos set under various conditions. It is worth mentioning that despite different settings her teaching style stays the same when the written-spoken ratio is taken into consideration. When it comes to the employment of *writing* mode in relation to the other investigated modes, Sample 3 and Sample 7 show almost the same level despite being representatives of considerably diverse groups. Furthermore, this tutor evinces also comparable levels of *gesture* usage under varied conditions. It indicates that modification of methods of production and corresponding higher employment of digital technologies does not change the individual presentational style as much as it might have been expected.

Finally, there is no doubt that the function of the teacher in the process of explication and instruction is fundamental; Bezemer et al. (2012: 12) note that the role of the teacher is becoming “one of the teacher as rhetor and designer of different sites as maximally effective environments for learning”. However preliminary, the results of this case study support this opinion; it seems that teachers are not being replaced by new digital technologies as it is sometimes suggested. On the contrary, teachers, or more generally educators, strengthen their position in the digital educational landscape by being designers of the effective environment for learning. By taking advantage of the multimodal expressive potential of online educational videos, they can increase communicative efficiency of teaching methods since using variety of modes caters to various learning styles and enhances the comprehensibility of the instruction. A further investigation based on the findings of the presented case study might help examine more closely the

instructional design of online educational videos with respect to multimodality and consequently improve performance of multimodal competence.

References

- Bateman, John A. and Florian Schmidt-Borcherding (2018) The Communicative Effectiveness of Education Videos: Towards an Empirically-Motivated Multimodal Account. *Multimodal Technologies and Interaction* 2, 59.
- Beege Maik et al (2017) Look into my eyes! Exploring the effect of addressing in educational videos. *Learning and Instruction* 49, 113–120.
- Bezemer Jeff et al. (2012) Using a Social Semiotic Approach to Multimodality: Researching Learning in Schools, Museums and Hospitals. NCRM Working Paper. NCRM (Unpublished)
- Bezemer Jeff and Kress Gunther (2015) *Multimodality, Learning and Communication: A social semiotic frame*. London and New York, Routledge.
- Chorianopoulos, Konstantinos; Giannakos, Michail (2013) Usability Design for Video Lectures. In Proceedings of the 11th European Conference on Interactive TV and Video, Como, Italy, 24–26 June 2013; ACM: Como, Italy, 2013, 163–164.
- Crook, Charles, Schofield, Louise (2017) The video lecture. *The Internet and Higher Education* 34, 56–64.
- Danielsson, Kristina (2016) Modes and meaning in the classroom – The role of different semiotic resources to convey meaning in science classrooms. *Linguistics and Education* 35, 88–99.
- Fiorella, Logan and Richard E. Mayer (2015) Effects of observing the instructor draw diagrams on learning from multimedia messages. *Journal of Educational Psychology* 108(4), 528–546.
- Fiorella, Logan et al (2018) Instructor Presence in Video Lectures: The Role of Dynamic Drawings, Eye Contact, and Instructor Visibility. *Journal of Educational Psychology*.
- Jewitt, Carey (2009) *The Routledge Handbook of Multimodal Analysis*. London: Routledge.
- Jewitt, Carey (2016) Multimodal analysis. In: Georgakopoulou, Alexandra and Tereza Spilioti (eds.) *The Routledge Handbook of Language and Digital Communication*. London and New York: Routledge, 69–84.
- Kizilcec René F. et al. (2014) Showing face in video instruction: effects on information retention, visual attention, and affect. Proceeding CHI '14 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Toronto, Ontario, Canada, 2095–2102.
- Kizilcec René F. et al. (2015) The instructor's face in video instruction: evidence from two large-scale field studies. *Journal of Educational Psychology* 107(3), 724–739.
- Kress, Gunther (1998) Visual and verbal modes of representation in electronically mediated communication: the potentials of new forms of text. In: I. Snyder (ed.) *Page to Screen: Taking Literacy into the Electronic Era*. London and New York: Routledge, 53–79.
- Kress Gunther et al. (2001) *Multimodal Teaching and Learning: The Rhetorics of the Science Classroom*. London and New York: Continuum.
- Kress, Gunther (2003) *Literacy in the New Media Age*. London: Routledge.
- Kress Gunther (2009) What is mode? In: Jewitt, Carey (ed.) *The Routledge Handbook of Multimodal Analysis*. London: Routledge, 54–67.
- Kruger Jan-Louis and Doherty Stephen (2016) Measuring cognitive load in the presence of educational video: Towards a multimodal methodology. *Australasian Journal of Educational Technology* 32(6), 19–31.

- Morell, Teresa (2018) Multimodal competence and effective interactive lecturing. *System* 77, Special Issue: Multimodal Perspectives on English Language Teaching in Higher Education, 70–79.
- O'Halloran, Kay L. (2005) *Mathematical Discourse: Language, Symbolism and Visual Images*. London and New York: Continuum.
- O'Halloran, Kay L. (2015) The language of learning mathematics: A multimodal perspective. *The Journal of Mathematical Behavior* 40, 63–74.
- Norris Sigrid (2004) *Analyzing Multimodal Interaction: A Methodological Framework*. New York and London: Routledge.
- Norris, Sigrid (2013) What is a mode? Smell, olfactory perception, and the notion of mode in multimodal mediated theory. *Multimodal Communication* 2(2), 155–169.
- Stull Andrew T. et al. (2018) Using transparent whiteboards to boost learning from online STEM lectures. *Computers & Education* 120, 146–159.
- Tang, Kok-Sing (2013) Instantiation of multimodal semiotic systems in science classroom discourse. *Language Sciences* 37, 22–35.
- Tang, Kok-Sing et al. (2014) An Integrative Framework for the Analysis of Multiple and Multimodal Representations for Meaning-Making in Science Education. *Science Education* 98, 305–326.
- Van den Broek Gesa S.E. et al. (2014) Effects of text modality in multimedia presentations on written and oral performance. *Journal of Computer Assisted Learning* 30, 438–449.
- Wang Jiahui and Antonenko Pavlo D. (2017) Instructor presence in instructional video: Effects on visual attention, recall, and perceived learning. *Computers in Human Behavior*, 71, 79–89.

Links

- Sample 1 (Group A) <https://www.youtube.com/watch?v=tVzaX9u6YAE&t=1425s>
- Sample 2 (Group B) <https://www.youtube.com/watch?v=HaOHUfymysuk&t=103s>
- Sample 3 (Group B) <https://www.youtube.com/watch?v=dPrzLjwxsOY>
- Sample 4 (Group C) <https://www.youtube.com/watch?v=FasJS4mmsn8>
- Sample 5 (Group C) https://www.youtube.com/watch?v=7kwpGDH_eUQ
- Sample 6 (Group D) <https://www.youtube.com/watch?v=Et4Y41ZNyao>
- Sample 7 (Group E) <https://www.youtube.com/watch?v=p8yesVjNdzY>
- Sample 8 (Group E) https://www.youtube.com/watch?v=Y-8tP0Pa_4&t=601s

SIMONA KORYTÁŘOVÁ studied mathematics and chemistry at the Faculty of Sciences of Palacký University. Later she studied English language and literature for secondary schools at the Faculty of Arts of Ostrava University, where she also obtained her doctoral degree in the field of English Philology. Her dissertation draws on Systemic Functional Linguistics and deals with a genre and multimodal analysis of educational video tutorials on infinitesimal calculus available at YouTube. She has been teaching English language for more than thirty years.

Address: Mgr. Simona Korytářová, Ph.D., Pod Zámečkem 3355, 73801 Frýdek-Místek.
[email: simona.korytarova@seznam.cz]



This work can be used in accordance with the Creative Commons BY-NC-ND 4.0 International license terms and conditions (<https://creativecommons.org/licenses/by-nc-nd/4.0/legalcode>). This does not apply to works or elements (such as image or photographs) that are used in the work under a contractual license or exception or limitation to relevant rights.

