METHODOLOGY

Open Access

CrossMark

"Peeling an onion": layering as a methodology to promote embodied perspectives in video analysis

Liv Kondrup Kristensen[1,2]

Correspondence: likr@pha.dk
[1]Center for social Education, University College Absalon, Roskilde, Denmark
[2]Department of Health Science and Technology, Aalborg University, Aalborg, Denmark

Abstract

This article proposes a methodological framework for analyzing video by adopting an embodied perspective. In order to deal with researching the complexity of human interaction that has been captured on video, structured ways for analysis are needed. In this article, the metaphor of an *onion* is used to conceptualize the process of unpacking the layers of observed interactions on video. Four different layers are identified: Foregrounding bodies, considering talk in combination with body, including the environment, and depth and adjustment through participant perspectives. To illustrate the process of analysis through this methodological approach, a worked example of video observations featuring classroom interaction is presented. While analysis of video through the step-by-step process in four layers is laborious, it is forcing the researcher to break with the habit of privileging talk as the base-line for analysis, sensitizing the analytical process towards non-verbal dimensions of interaction, while bringing in material dimensions, as well as the voices of participants in order to understand embodied interaction as situated activity.

Keywords: Methodology, Video analysis, Embodiment, Classroom research

Introduction

Video, as a method for data collection and analysis in classroom research, can be a powerful tool that allows the capture and review of interactions in detail, and it can provide more authentic insights than, for example field notes Knoblauch et al. 2012. Video enables the capture of details that a one-time observation might miss, so data can be reviewed multiple times, or allow for access to non-verbal matters, such as facial expressions, intonation or stance-taking (Cohen et al., 2011; Derry et al. 2010; Knoblauch 2009). These obvious benefits of video have prompted researchers to make use of this medium in a wide range of ways. Video has been used to document classroom practices (Clarke et al., 2006), create accounts or narratives of teacher (Cowie et al., 2012) and student practices (Radinsky 2007), create visual platforms for discourse analysis (Goodwin et al., 2012), and expand ways for analyzing interactions, to allow for a focus that is more than just talk. Examples of the latter are, including body posture, gesture and facial expressions (Streeck 2014) or emotions (Ritchie & Newlands 2016). Video can also function as a medium for sharing and communicating findings in ways different from more traditional transcription conventions, e.g. the Jefferson Transcription System (Jefferson, 1984). A

very specific appropriation of this approach is for example found in the works of Norris (2012), who presents video data in a way that enables the readers to not only gain a visual impression of the interactions, but also temporal aspects by using consecutive still-frames from video combined with talk in multimodal transcripts.

Delimiting the field

Despite the advantages of video, the method poses a number of challenges that touch upon issues ranging from the naturalness of the data (Schnettler & Raab 2008), how technology manipulates and frames the data obtained (Laurier & Philo 2012), to methodological questions about selection, analysis (Derry et al. 2010), and transcription practices (Knoblauch et al., 2012a, b). In terms of transcription, researchers in social sciences may face rich, complex and/or large quantities of data, and need to reduce the visual, kinesthetic, and acoustic information into simple transcripts that are fit for communicating to wider audiences, often in text formats that suit journal requirements. As such, talk is often the primary type of data presented, supplemented by still frames from the video to provide context to the discussion. In terms of capturing and retaining interaction, what is won is at the same time lost in respect of transcribing and communicating, as much of the richness is omitted in favor of clarity. This means that bodies and materiality may succumb to being presented as auxiliary aspects to talk, and this may be grounded in two possible reasons. On the one hand, popular mediums for communication research, such as journals and books, are customized to fit written communication. Hence, scholars who want to publish must write. On the other hand, there is a history and culture of communication by writing in the scholarly community (Ayass 2015; Schnettler & Raab 2008; Soeffner 2012). Therefore, the written language is often the preferred medium that we have shaped our world around, specifically in academia. With the availability of video data, this norm for communicating research with its insistence on rich and detailed data, calls for the researcher to find new and innovative ways of working with and communicating the observed (Knoblauch 2009, 2012).

The article is proposing an analytical framework for working systematically with video. This framework attempts to provide structure for a researcher to become more sensitive and attentive to tacit, embodied, material or unspoken dimensions of video data (Polanyi 2009). Dealing with the complexity of video data has been picked up by others in the field. For example Derry et al. (2010) formulate an overarching strategy for focusing the analysis on three central steps of *indexing, macrolevel and microlevel coding/ narrative summaries/ diagrams,* and *transcription* that help the researcher understand and interpret interactions. The general tenets of the above approach is also reflected in the works of Erickson (2006) and Knoblauch, 2009, 2012, Knoblauch et al 2015). Erickson (2006) suggests the use of context analysis, microethnographic discourse analysis, and conversation analysis as an inductive approach, where verbal and non-verbal activity is considered equal. Erickson suggests a 6-step approach, where the researcher begins by *reviewing* video multiple times, to then move onto detailed *transcription* of one event, before reviewing with others and looking across the video material to determine typicality. This approach seeks to level verbal and non-verbal activity in its insistence on noting both types of interaction in each step, for example by using a horizontal chart to avoid overemphasizing talk. Knoblauch, 2009, 2012, Knoblauch et al. 2015), suggest a cyclical research

process of *indexing – selecting – detailed transcription – comparison*. The process enables the researcher to gain an overview, determine relevant sequences, demarcating sequences, and comparing, which again informs new cycles of analysis. Yet, Knoblauch et al. (2015) also consider how to structure the process of transcription and analysis. They identify several different layers, which are then unpacked, starting with talk and prosodic features, moving on to non-linguistic features, such as interactions with actors (can also be e.g. touching objects), movement towards the actor they are interacting with, and gaze. Observations are aligned in flow diagrams, expressing sequentiality and simultaneity (for example see Knoblauch et al., 2015, p. 109). Horizontal charts are not the only way of structured video analysis. Goodwin (2007, 2009) adds rigor into video analysis by developing highly detailed transcripts in which talk and prosodic features are connected with an analysis of posture/gaze/gesture of the embodied participants and features of the environment that are given relevance by the participants. By embedding drawings into the text, he shows how participants position themselves at a particular moment. Goodwin's approach to structured analysis is highly detailed and works best for smaller amounts of video data. Larger amounts of data call for different approaches. In working with video data form large international studies, Klette (2009) applies theory-driven coding to structure analysis, which is very different from the above mentioned approaches. Using pre-defined categories and sub-categories, the course of e.g. a lesson is mapped, and cross connections to other data sets are being made. The value of quantitative approaches for systematically unpacking video is likewise emphasized in Jacobs et al. (1999), who look at quantitative analysis as supplement and validation of an initial qualitative analysis. They argue for a cyclical process of 6 steps: 1) watch/discuss, 2) generate hypothesis, 3) develop code, 4) apply code, 5) analyze/interpret, and 6) link to video, before going back to watching and discussing again. All these methods for video analysis suggest certain ways and sequences to strengthen the outcomes generated from complex video data. The interest in the method suggested here is in the foregrounding of embodied actions.

In what follows, I will first outline the embodied perspective that is integral to this article and its significance to video analysis. The discussion on bodies as the site of investigation and its inherent challenges propels the proposition of "layering" as a methodology for analyzing video. This methodology is presented using the metaphor of an onion, and is afterwards unpacked using video material from a research project looking at classroom interaction.

Why adopt an embodied perspective on video research?
Human perception is not solely a product of the mind but is rooted in the body, which is the anchor for being in the world (Merleau-Ponty 2012). This means that the very materiality of the body shapes perception. The physical position and shape of the body, for example, determines what we can perceive. The experiences of the world are embedded in the body as repertoires of action that shape the mastery of everyday life, for example jumping over a stream, writing with a pen, or tying a shoelace (Merleau-Ponty 2012; Thøgersen 2014). Acting in the world is therefore an expression of people's perception and experience of the world by means of their bodies, and as such we "*are* our bodies, and bodies *are* lived experiences" (Zembylas 2007, p. 21). For the purposes of the argument in this article, embodiment has two consequences for video analysis.

Firstly, interactions (embodied practices) cannot be reduced to talk, gesture, or posture as expressions of superior cognitive processes. These actions are fundamentally rooted in the lives of human beings, and so they must be seen as such, bearing with them significance and meaning in the concrete social situation (Goffman 1959). Consequently, social interaction is about lived bodies making sense of the world, and doing so by means of the body with all its experiences, habits, and emotions, shaping the way in which the action is performed. Understanding the meaning and signification of an action is therefore to also know the person and their (embodied) incentives to act in certain ways (Goffman 1959). We thus have to look behind what lies at face-value (gestures, talk, posture, facial expressions, and so forth) and consider an emic perspective; looking at a system of interaction from the position of the insider, creating accounts that are truthful in the eyes of the participant (Knoblauch 2009).

Secondly, acknowledging embodiment as acts of sense-making opens up a view of the body as communicative. By being and acting in the world, embodied acts communicate about how a person makes sense of, and relates to the world. As such, an act like a gesture is not merely a tool for supporting talk, it is simultaneously an expression of past experiences and a way of making sense of the given situation and environment that can be picked up and interpreted by others. However, the body as communicative is not a static phenomenon. According to O'Loughlin (1998), the communicative body is essentially "a body *in process* of creating itself" (p. 279), which means that the body and its emergence can never be fully captured or understood, as it is continuously growing and developing through its immersion and engagement with the world. Different theoretical perspectives foreground certain qualities of the body, and neglect others in doing so. As such theory can, "neither describe nor prescribe such a body; all that can be effected is the bringing together of fragments of its emergence" (O'Loughlin 1998, p. 279). The notion that the body cannot be captured by a single theory is central to working with video. Video generates multiple forms of data (picture, sound, text, and so forth), which can be considered as 'slices of data' (gesture, prosody, talk, posture, facial expressions) (Glaser & Strauss 1967). To make an example, examining gestures grants a particular vantage point from which to understand a social phenomenon at hand. A gesture can, by means of still-frames, drawings, short video segments, or detailed descriptions, be foregrounded and placed in the context of a social interaction, where sequential analysis might reveal something about the relation between that gesture and prior/later acts. However, paying attention to gestures also means examining only a fragment of the emergence of the body. The argument is thus made that researchers need to take care and consider which fragments are foregrounded in an analysis, and how these fragments are represented in transcripts and dissemination products.

These theoretical premises are the guiding principles of the following sections that consider an (embodied) methodology in more detail.

Methodological considerations

Foregrounding bodies

Methodological approaches in framing the data collection and analysis of video data, in relation to embodiment, has gained useful insights through phenomenological approaches (Streeck et al., 2011), especially when the body is regarded as a vehicle for

being in the world and understanding the ways of comporting ourselves (Dreyfuss 1991; Merleau-Ponty 2012). In multimodal analysis for instance, language is viewed as only one mode of communication which may or may not take up a central role in a given situation (Norris 2004). As such, at least theoretically, bodies and materiality are paralleled to talk since each mode carries interactional meaning for an individual, although Norris does comment on the issue that language often plays a central role in interaction (2004). Without employing the term 'multimodality,' Kendon (2004) similarly notes that it is unproductive for researchers to separate gestures from language as people, when they speak together in face to face situations, always mobilize several modes of communication. He states that:

> "Every single utterance using speech employs, in a completely integrated fashion, patterns of voicing and intonation, pausing, and rhytmicities, which are manifested not only audibly, but kinesically as well, and always, as part of this, there are movements of the eyes, the eyelids, the eyebrows, the brows, as well as the mouth" (as cited in Streeck et al., 2011, p. 9).

A methodology that details interactions and the role of the body includes Conversation Analysis (CA). Here the researcher considers a conversation to include at least two people: a speaker and a listener. To identify the roles of the two, body posture and body orientation are identified in addition a deep analysis of talk, for instance by including physiological markers like pitch and volume prosodic analysis (see, for example, Goodwin & Goodwin 2004). Despite the recognition of bodies in interaction there is a 'lingering dualism' that emerges when talk and text are the preferred medium for analysis and communication (Streeck 2003). Even when working with video, the verbal seems to maintain a focal position and often forms the baseline for understanding interaction (Knoblauch 2009). Since the whole body mediates the engagement with the world, non-verbal interaction with people and materials should be considered equally important to talk (Streeck et al., 2011).

   The idea of starting with visible comportment as the first step in video analysis may address the issue of the complexity in understanding human interactions, but it comes with a lack of orthography that could be used for the transcription of visual and tactile conduct (Luckmann 2012; Schnettler & Raab 2008). It seems the more interest is in the fine detailed styles of embodied action, the greater the need for specialized codes and notation scores. However, this brings with it having to address challenges of synchronicity and juxtaposition of different scores in combined transcripts (Luckmann 2012). Fragmenting embodied actions into different modes is also problematic from a theoretical perspective. Fragments might provide glimpses of the emergence of the body, but it is only when coming together that these fragments act as the vehicle through which meanings are expressed (O'Loughlin 1998). O'Loughlin suggests that a different approach to *recover* the body is by way of re-examination of emotion (1998). She notes that the body "*as* action and communication can only be so through emotion", and that emotion functions as "a guide to, and preparation of, the individual's social action" (O'Loughlin 1998, p. 279–280). This is in line with Merleau-Ponty's thinking, who understands emotions as practical consciousness that guides how people make sense of situations, and enables them to act and react in social interactions (1998). Hence, the

suggestion here is to look at 'affective' dimensions, noting the emotional qualities that can be identified through embodied performances so that they can become resources for a more holistic understanding of interactions.

Considering talk in combination with body

Foregrounding bodies does not mean neglecting talk, on the contrary, talk is just as essential for understanding elements of activity (Knoblauch 2009). Talk is a vehicle of human action (Schegloff 1991), and as such talk is corporeally intertwined with other forms of action, like gaze and gesture (Goodwin 1981) that are crucial resources when participants attempt to align themselves towards the activity of the moment (Goodwin 2009). Essentially, talk like visual orientation or gestures can be used to identify what is experienced as important in the context (Goodwin 2000), and in doing so expand and deepen understanding of the video segment. Utilizing talk as the base-line for analysis is an attractive option as the conventions for working with talk are well established and have been used across a number of research traditions (Peräkylä 2005). Talk is also the most common approach to the process of transcription since it provides a 'vehicle', 'resource' (Heath & Hindmarsh 2002), or 'location device' (Knoblauch 2009) that details participants' conduct and identifies sequences. As there is no general orthography for the transcription of visual and tactile conduct (Heath & Hindmarsh 2002), there is no established way of combining talk with visual/tactile behavior. Different approaches have been developed that incorporate elements of visual, audio, tactile, and environment to varying degrees. Heath et al. (2010a, b) expanded written transcripts by adding annotations of embodied actions and in doing so developed them into complex scores. This form of transcribing retains the sequentiality and action-turns, and includes embodied actions as far as they are relevant to the analysis (Knoblauch et al., 2015). Another form is notation in score form (see for example Raab & Tänzler 2012) where the data is divided into different modes that are then described in a score. While a notion score is able to retain much information about the embodied dimensions visible in the segment, it is difficult to move from a detailed coding to vernacular transcripts (Knoblauch 2009; Luckmann 2012). A third format can be found in the works of Norris (2004, 2012), who uses heuristic modeling when combining talk and images to communicate, not only her analysis about what is taking place, but also to add a sense of the temporal and embodied mode of interaction. She transcribes talk to written text and places it on top of still frames in near proximity to the speaker in a sequential manner to show when something is said in relation to the embodied act depicted in the image. The text is then manipulated to indicate rise in pitch and intonation by means of big and small fonts as well as word-art. In creating aggregates Norris makes the text part of the image, which breaks with the structural difference between text and image that feeds into the 'lingering dualism' of ascribing more importance to texts than images (Streeck 2003). The merger between text (audio) and image (visual) is furthermore accentuated by using heuristics to imbue words with emotion. As such, the visible styles of conduct, combined with text that give emphasis, provide a sense of the atmosphere of an interaction.

Including the environment

When participants interact, they draw on a wide range of social and material resources that are used to negotiate their lives (Streeck et al. 2011). Recognition of and attention

to the environment is therefore central to interpreting and understanding interaction. In an early study, Suchman (1987) demonstrated how the interpretation of people operating technical objects cannot be guided by normative rules. She asserted that a suitable methodology that is used to investigate the usage of objects (in her study: copy machines) represents accomplishments that are situated, contingent and interpretive. Not only is the use of technology highly contingent and situated, but technical instruments, such as surgical tools (Bezemer et al. 2014) or power points (Schnettler 2012), scaffold interaction as many of them have become an automatized part of reasoning. The consideration of the material environment is not without difficulties when interfering with the orders of mundane reasoning and interaction (Heath & Luff 2000). What can be gathered by examining materials and objects, as part of the environment where interactions take place, is that talk and embodied communication are situated within complex material settings that make participation simultaneously intelligible and coherent. The strong link between embodied action, structures in the environment, and talk is further emphasised in the works of Goodwin (2000, 2007), who showed how gesture and talk are environmentally coupled in everyday interactions like doing homework or playing hopscotch. Schmitt and Deppermann (2007) also explained the concept of "interaction space", how the interplay of physical circumstances with their particular features has implications for how interaction is structured, and also what is accomplished within those interactions. Furthermore, how these interaction spaces are connected to structures of relevance, which for example can be expressed through the symbolization of inclusion and exclusion (p. 96, as cited in Streeck et al. 2011, p. 11).

The challenge of taking note of and including the environment in an analysis relates to the question of what is relevant (Knoblauch 2012). Schegloff (1991) called this 'the criterion of relevance', which holds that what is relevant to the analyst must be shown to be relevant to the participants. Goodwin (2000) notes that this can be observed by looking at the visible orientation of the participants as a spotlight onto those features in the contexts that are important and relevant. There are different resources available when trying to discern what is relevant, such as talk, gaze or gestures. These resources can be foregrounded in different ways in transcripts. Goodwin for example, uses arrows to indicate the direction of gaze, emphasizes certain words in bold or italics to indicate particular qualities in talk, and includes drawings of the participants bodies and immediate shared objects in the environment as ways of making visible that which is relevant to their interaction (2000).

Validation by emic perspectives through participant voices

A focus on relevance calls for intimate knowledge of the field and video data (Knoblauch 2009, 2012), which transgress the information that can be extracted from the video material alone. Knowing the field and interpreting the interactions that unfold on the screen entails understanding the culture in which the interactions unfold. It is the very situated feature of practice (that is habitualized, routinized, and institutionalized in relation to a particular environment, group, and context) that makes it difficult to gain access by means of video (Knoblauch 2009). Knoblauch argues that "As objectified as the meanings might be, they are always related to someone who needs to understand them" (2009, p. 185). He proposes ethnography as a way forward to clarify the meaning and signification of visual

elements of recorded data (2012). Ethnography offers insights into the context in ways that video observation does not by means of 'reflexivity'. Reflexivity means that participants 'frame' or 'indicate' how their action is to be understood, as opposed to just acting, and in order to understand the basic intent behind an action knowing the culture becomes central. Ethnographic knowledge becomes particularly important when the spoken is less significant, as the interpretation of the visual is dependent on knowledge of the context in which they unfold, and the participants involved (Knoblauch 2009). Participant observation and video-stimulated recall interviews (Morgan 2007; Raingruber 2003) are methods for retrieving the reflections and subjective perspectives of participants who are at the center of the researcher's attention. Addressing an emic perspective by examining the participants' point of view enables questioning how matters of identity, history, and culture come to shape what can be witnessed on video. The emic perspective thus becomes a way of supporting or adjusting the interpretations of embodied actions.

In the following, I utilize the above considerations together with the metaphor of *the onion* to formulate the idea of *layering* as a methodology for video analysis.

Layering: Using *the onion* as metaphor

The methodological considerations above were presented as four different dimensions:

1) Foregrounding bodies – the visible layer
2) Considering talk in combination with body – the audible layer
3) Including the environment – the material layer
4) Depth and adjustment through participant perspectives – the emic layer

These dimensions are here conceptualised as four *layers* that can be found in video data, where each layer represents a different vantage point from which to understand embodied human activity. The main argument of this article is that if embodied activity is to be understood holistically, the four layers need to be brought together. As shown in the above, the bringing together of different layers is fraught with difficulty, where embodied aspects may lose their significance. To get around this issue, the metaphor of the onion is adopted to give structure to the bringing together of layers. An onion is an organic entity; it has no pit, but consists of multiple layers. As such, each layer is an essential part of how we come to understand the onion. Much the same counts for the above-mentioned layers – each layer is part of the 'array of affordances' (Hutchby 2003) video makes available, which conditions how we come to interpret and understand an event.

The idea of *the onion* as a metaphor for the methodology of layering is visualized below in Fig. 1:

In this article, *the onion* metaphor represents a particular approach to video data, where the process of 'peeling back layers' simultaneously is an act of bringing together the layers. I envision it like this: To examine the onion, the layers must be layed open and peeling back one layer at the time achieves this. When peeling back one layer at the time, what becomes important is that the second layer peeled back, is not understood in isolation from the first layer. On the contrary, the way in which we come to talk about and understand the second layer (in this article, talk), is shaped by what was brought forth in the first layer (the visual). Hence, each separate layer does not only provide a new vantage point, but it adds *depth* to the growing interpretation. The

Fig. 1 Illustration of the metaphor of *the onion* as a methodological framework
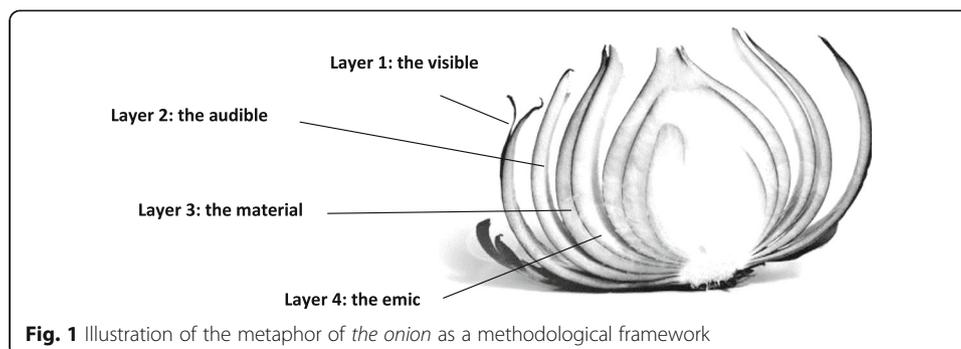
metaphor of *the onion* therefore enables analytical considerations about not only which layers that are examined, but also the order in which they are examined. To exemplify this approach, the following sections show the structured analysis of video data.

The data

The data used in this example was collected in the spring of 2014 as part of a project that investigated pedagogy of embodiment in physics education in upper-primary school. The study was conducted in a Danish primary school that formally adopted a policy of integrating movement and exercise into all subjects. The data collection included approximately 8 h of video-recorded participant observations over a period of one month, digital photographs, student and teacher work samples, field notes, interviews with the teacher before and after each class, group interviews with the students, and video stimulated recall interviews with selected students. In this article the video-observations, field notes, group interviews, and video-stimulated recall interviews are used.

The analysis shared here is a short video segment of 45 s that shows the interactions of a group of year 8 students working with a classroom activity about the Doppler Effect.[1] In this episode, the students just entered the hallway and have started a task they received from their teacher. The task was described on paper and asked that each group member should take a turn at running down the hallway while carrying a device (mobile phone) that produced a constant high pitch tune. The remainder of the group had to stay in the hallway and listen to the sound as the runner approached, was close and then passed them. This would enable the listening group to evaluate how the quality of the sound changed.

The reason for selecting this particular episode was to provide an illustration of how a 'layered' approach both deals with and emphasizes the complexity inherent in human interaction, which in this particular episode was the process of negotiating running in the hallway. A layered approach to this episode enables the recognition of this event, as a complex intertwining of feelings towards the task, social relations, habits, and perception of physical capital in self and others. In the analysis that follows, I exemplify the unpacking of this episode in four steps: layer one, the visible layer with focus on the visible aspects of the student's interaction. This means noting how they utilize their bodies to position themselves towards the task and in what manner. This first analysis was achieved by watching the video without sound to become sensitive to visible interaction. The analysis of layer one, provided the canvas for layer two, the audible dimensions of interactions. This time the

video was watched with sound, taking note of talk in the visual context, including para-verbal features specifically pitch to focus on emphasis placed through verbal interactions. Layer three, the environment, examined the physical-material environment in connection with the students' interactions to get a sense of habitation in connection with talk and movement. Finally, the analysis of layer four, the emic perspective, required realigning the researcher interpretations with the students' perspectives to bring in students' voices and provide more depth to the interpretation of video.

Layer 1: Foregrounding bodies – The visible layer

When watching the video muted, it takes on an almost unaccommodating character, as the unfolding interactions seem deficit or unfinished without the clues that sound usually provides us with. Yet, putting that feeling aside what first springs to mind (in the case of the author), is the everyday and mundane character of the episode. Students move around in ways that appear accustomed and accepted in the social space by the calmness and ease of their reactions.

Please click the following link to view the video clip without sound to make your own assessment:

Embodied Pedagogy Clip B: (https://youtu.be/OvBpUQlS5Pg).

It is perhaps the very naturalness and ease of their movements that is at the core of what makes it difficult to discern the complex orchestration of different modes of movement (gesture, posture, facial expression, travelling) from each other in locating a particular reference point, which could serve as a location device. One way to approach this challenge is to locate the necessary vocabulary for talking about movement in ways that relate movement to interaction. One such language for talking about embodied interaction and style of conduct is found in the Laban Movement Analysis (LMA), developed by Rudolph Laban (1975). The system provides a theoretical and experiential system for the observation, description, and interpretation of human movement (Laban & Lawrence 1974). The analysis of the visible layer focuses on three major categories of movement elements as defined in LMA: Space, Shape, and Effort (for overview see Konie 2011). *Space* explores where in space movement takes place, and how movement relates to the kine-sphere.[2] *Shape* is about form and forming, and explores how movement travels into space and creates shapes. *Effort* is about how we move and how certain movements are accomplished and with what energy (direct/indirect, strong/light, quick/sustained, bound/free). These categories, their related adjectives and qualities, makes it possible to describe movement that can be witnessed in video across time and space, while also being sensitive to the expressive and affective stance inherent in the different movements.

Following repeated viewing cycles of the above video building on the LMA categories, significant events were captured in still frames that functioned as the reference point or canvas for the analysis.

Students negotiating running – Space, shape, and effort

The video shows a group of students in the school hallway. The group consisted of four students, Alfons, Mira, Hai and Adi (all pseudonyms). In what follows, I use framegrabs to illustrate how the students negotiate the task through movement.

In the first frame (Fig. 2), three of the four students in the group are visible. They are conferring with a boy from another group, who has already completed the task. Mira,

Fig. 2 Alfons and the group attentive to a classmate telling and showing them the outline of the experiment

Alfons and Hai remain put while the boy moves animatedly, pointing in the direction of the far end of the hallway, and suddenly hauling his arm to point in the other direction. While he is making a demonstration, Mira is sitting on the ground, while Alfons and Hai are leaning against the wall in a slight hunching position. Despite the difference in levels (low, mid, high), their bodies are aligned. They are all positioned horizontally with their backs against the wall, but also in terms of the *shape* of their body, where they all to some degree create a ball shape through hunched forward shoulders and arms closing or folding in front of themselves. The ball shape is characterized by shrinking of the internal kinesphere, shortening their vertical horizon, narrowing their horizontal dimension, and hallowing their sagittal dimension. The shape quality associated with this kind of motion can be described as sinking, enclosing, and retreating. The lack of tonus in their bodies, provides, in terms of *effort,* their movement with a heavy weight and bound flow as their movements are controlled and contained, and can be stopped at any time. They almost seem drawn towards the floor and indirect in their embodied presence and attention to what happens around them, this is in contrast to their gaze, which is shifting around and their verbal activity, showing stability and presence.

25 s into the video, the boy leaves and Alfons changes his position. With his back against the wall, he slides down to a seated position next to Mira (see Fig. 3). His back is pressed against the wall and his legs are pulled towards his body to a lesser extent than Mira (see Fig. 4). This action adds to the heavy weight quality of his movements, and together with Mira they project a lack of willingness to use their bodies.

Fig. 3 Alfons in the process of sitting down

Fig. 4 Alfons sitting down

Fig. 5 Different positions adopted

By sitting down, Mira and Alfons utilize the *space* in a different manner than Hai and Adi, who are still standing up. By using the space differently they subsequently also appear to adopt a different, more reserved stance towards the task. Their stance becomes more pronounced as Adi joins the group 32 s into the video (see Fig. 5). Adi is moving across the hallway, his body is straight, making a pin shape and in doing so growing and lengthening across the general and vertical dimension. The shape associated with his movement is rising, spreading, and advancing. His efforts can be characterized as free flow, having a light and active weigh, yet also directing and assertive.

Albeit descriptive, layer one with its use of effort, space and shape descriptions provides a sense of the manner and style in which the students approach the task. What becomes evident through the descriptions are the different stances to the task of running, not only in terms of how the students are positioned in the hallway, but also the manner in which they position themselves. In what follows, this analysis is deepened by attending to talk.

Layer 2: Talk in combination with body – The audible layer
This layer seeks to add depth to the above analysis by considering talk. As the first (visible) layer constitute the basis for the analysis, I have looked to Norris (2012) to find ways to combine captured interaction (still frames) and talk in aggregates. In her

multimodal transcripts, Norris places excerpts of talk onto still frames. The text does not only explain the action, it is furthermore placed in a particular way ('before', 'on top' or 'after' the participant) to indicate temporality and sequentiality. The text is also manipulated using Word-Art (adding shape and font) to indicate intonation and force.

Negotiating running through talk and interacting bodies

When watching the episode with sound, the first encounter with the task of running is embodied by the boy, who explains by means of his body and voice how they are supposed to run. He supports this explanation with the sound "schiiiiiuuung", denoting a fast movement from right to left (see Fig. 6). When he is about to leave, Mira asks who should run. She continues by suggesting Alfons as a runner, but quickly discards this idea by pointing to Alfons' disinclination to running. As an answer to who should run the boy from the other group answers "somebody fast" emphasizing the word 'fast' before moving away from the group. This is the first time that the idea of being *fast* is introduced and verbalized as something of value to the task. Mira reacts to this by calling out that Adi should run, and in doing so positioned Adi as a capable and fast runner. Fig. 6 shows, the conversation as it evolved over time, visualized through the text going across the images. Examining talk afforded new dimensions in the embodied interactions to be noted and expanded on, and thus resulted in the identification and addition of new still frames to the existing canvas. Please also click on the following link to view the video with sound, to note how the student use intonation and volume to emphasize their stance towards running. In the video, the students speak Danish, the text superimposed on the still frames is the English translation:

Embodied Pedagogy Clip A: (https://youtu.be/ecNysU40nhc).

Mira continues, exclaiming that she does not feel like running, which is supported by Alfons, who also states that he does not feel like running either, which he reinforces by sitting down next to her (see Fig. 7).

Fig. 6 Boy explaining how the group should run

Fig. 7 Mira and Alfons taking a stance
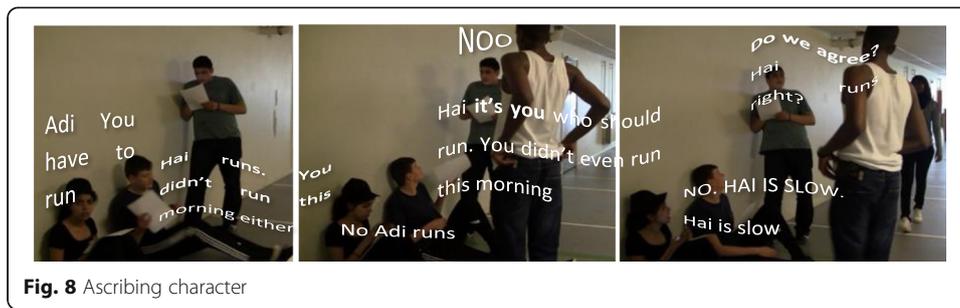
Fig. 8 Ascribing character

Mira and Alfons placed themselves in a position that emphasized when they explained to the others that they did not want to run. As a strategy to avoid running, Mira and Alfons draws the attention to Adi and Hai by means of talk, gaze, and gesture (Fig. 8). Mira explains to the others that Adi has to run, while. Alfons wants Hai to run. Mira maintains that Adi should run, but Adi says 'no', but argues that Hai is too slow and not right for a running task. Alfons and Adi still want Hai to run and Adi steps in and closes the argument by saying "do we agree? Hai runs right?"

By adding talk to the still frames, the movement qualities highlighted in layer one are contextualized. Thus from the analysis we see that the retreating and enclosing movement adopted by Mira and Alfons is part of the process of negotiating the task of running. Their stance not only underlines their unwillingness to run, but remaining in a seated position, seems to exempt them from further discussion about who should run and whether they are 'qualified' to run. In the next layer, I try to understand these positions on the backdrop of the environment they are positioned in.

Layer 3: Including the environment – The material layer

Physical structures and objects matters to the participants because "man is reliant on outside-the-skin control mechanisms for ordering behaviour" (Geertz 1973, p. 44). Materials including seemingly mundane objects and structures shape people's expectations on what to expect in such an environment and become part of epistemic configurations (Roehl 2012). Layer three examines physical objects and structures that are visible in the video as resources for understanding the interactions between the students.

The hallway

The hallway is a long space that connects the main building with a multi-purpose hall, toilets, and three specialised classrooms. Figure 9 shows the hallway is a place for storage,

Fig. 9 Decorations on the wall and the floor in the hallway

Fig. 10 The school entrance

but there are no chairs, tables, or other furniture suitable for recreation/working available to the students. The hallway is a space for commuting between learning spaces.

The hallway includes different objects, pictures of athletes being physically active, such as the soccer player opposite Adi in Fig. 9. The materiality of the hallway is also shaped by the lines on the floor. The lines are indicators of a running track and there are markings for every five meters throughout the entire hallway. The lines continue outside the school and connect with the outline for a running track at the entrance of the school as shown in Fig. 10.

Embodying the hallway

The hallway is different to objects in the classroom that that organize (Geertz 1973) learning together. Since the hallway does not include equipment such as tables or lab stations that suggest collaboration, the students draw on the resources that are present in the hallway (or lack thereof) to shape their behavior. In the above analysis, the students' discuss running as something that needs to be fast. When trying to understand this idea of running as fast, resources such as the lines in the hallway or the pictures of the athletes provide the impression of an environment that denotes certain performative ideals embodied by the world of sports.

In the above analysis, Mira and Alfons' behavior (talk and movement) was interpreted as retreating from the task of running. Yet, when considering the lack of tables and chairs in the environment, their behavior can be seen in a different light. When Mira and Alfons are sitting down they have their assignment sheet resting on their legs, as if using their legs as table. Hence, their seated position can also be interpreted as a position that allowed them to focus on the text. Such an observation is not at odds with the prior interpretation, but instead nuances the understanding that their actions perhaps also are motivated by practical needs rather than only unwillingness to run. In the next layer these interpretation are weighted against the personal narratives of the students.

Layer 4: Depth and adjustment through participant perspectives – The emic layer

Layer four examines the voices of the students that are embodying the environment. By asking students to re-narrate their experience of the situation through video-stimulated recall interviews (Morgan 2007; Raingruber 2003) additional insights were gained into the sentiments behind certain actions. This also allowed realignment of the interpretations of observed actions by drawing on these interviews in combination with the findings from the first three layers. The participants' perspectives were compared with the existing transcripts; this was to bring in personal voices that inform how certain actions

came to make sense in the activity. Mapping the stories onto the existing canvas of findings was also done in an effort to create transparency between the participants and the researcher in the process of analyzing and interpreting.

Participants re-narrating their activities

In Figs. 11 and 12 below, meaning condensation of interviews (Kvale & Brinkmann 2008) are combined in aggregates with the existing transcripts.

Through an emic perspective new perspectives on the task of running emerges, that were not previously accessible. Mira raises the feeling of insecurity when being filmed by her peers as an argument for opting out of running, while Alfons notes that their actions of sitting down are also a product of how they would usually act in the hallway during recess. Yet it also shows that what could be construed as a hard tone towards Hai, from the perspective of Alfons is experienced as friendly bantering. Thus, examining interactions through the voices of the participants helped to adjust and deepen the overall findings, to show for example, how a retreating act of sitting down can explained by different sentiments, and open up for new understandings of the actions.
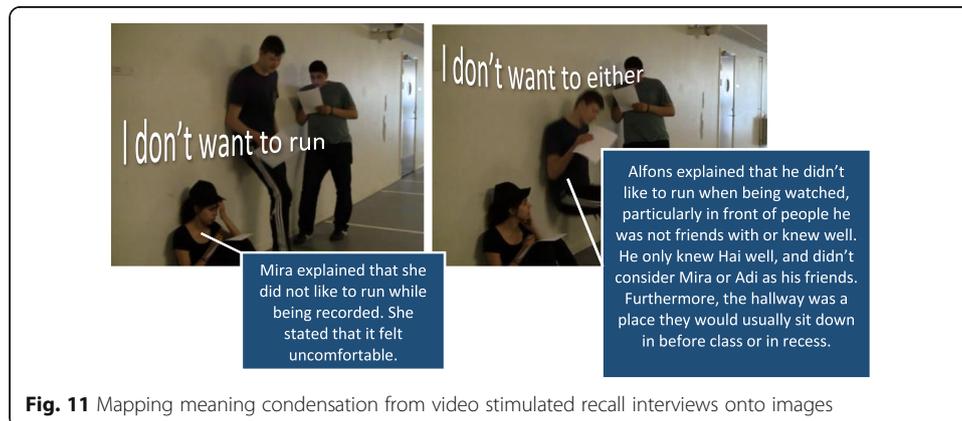
Fig. 11 Mapping meaning condensation from video stimulated recall interviews onto images
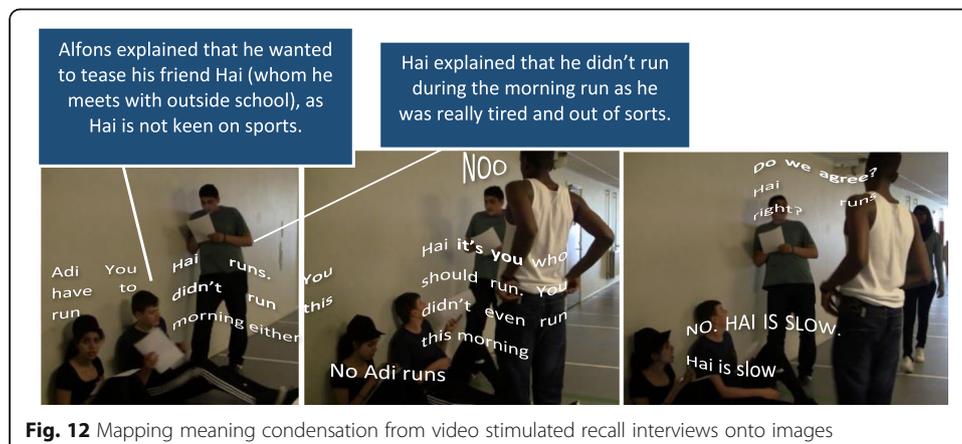
Fig. 12 Mapping meaning condensation from video stimulated recall interviews onto images

Discussion

This paper set out to propose a structured framework for analyzing video with a particular emphasis on highlighting embodied dimensions of interaction. Inspired by the metaphor of the onion, the result was a *layered* approach that advocate for a process of systematically peeling and merging layers to qualify and substantiate interpretation and understanding when working with video. The approach conceptualizes video data as consisting of multiple 'slices of data' (Glaser & Strauss 1967), and argues for an approach where data slices are disassembled to be reassembled. This procedure is suggested to heighten video researchers' sensitivity to work with video beyond focusing on talk, and muting the video, focusing on embodied performance, looking for pitch or gesture will aid a researcher to accomplish this. It is the process of analyzing and re-layering that is often difficult (Knoblauch et al. 2012; Luckmann 2012) and this is where this methodological approach aims to make a contribution. Qualitative video research that is aiming to include different modes of the body into analysis of human interaction such as Bezemer, Cope, Kress and Kneebone (2014) and Heath and Hindmarsh (2002) can add depth to the interpretations of human interaction. Identifying insights gained through video by using multimodal ways of analysis (Norris 2012) or notations of embodiment (Goodwin 2007) addresses the complexity of conveying what has been identified, but many times a dominant focus is on talk as the primary location device (Knoblauch 2009). While this as argued above can be, a way to reduce the inherent complexity of video data, at the same time it presents a process of re-layering with the need to arrange other modes to fit the structural organization of talk (Bezemer & Mavers 2011).

While conventions such as the Jeffersonian notation system (Jefferson 1984) provide high level of details through text, other modes offer additional depths. The methodology presented in this paper takes a different approach by privileging the body and its material and environmental encounters in the analysis. This is accomplished by looking at movement before talk, which opens up for the possibility to build a transcript around different and more holistic ways of representing the body. Streeck (2003) noted how, in privileging talk, there exists a 'lingering dualism', even within video analysis. I wonder if privileging the body before talk could be characterised as 'dualism in reverse'. However, analysing video always has a starting point, a modus operandi with which to approach the data. When removing talk, a more ambiguous form of communication becomes the basis for interpretation. It is ambiguous because individual modes such as gestures, gaze, or posture in themselves do not afford meaning, as people always mobilize several modes of communication simultaneously (Kendon 2004). Yet building on LMA in the first layer, broader descriptions of movement were enabled, which were not 'mechanical', but had a more emotional and qualitative character. These affective descriptions of movement, such as the Mira's 'retreating', 'introverted' and 'passive' posture, describes more than a given movement. It ascribes a character and manner with which the action was carried out, and in doing so, affords a more holistic approach to the visible properties of interaction than when the body is reduced to a posture, a gaze or a movement of the arm (Goodwin 2000, 2007, 2009) or the duration and coordination of intracorporeal actions (cut, lift, grasp and so forth) (Bezemer et al. 2014). Watching the video with sound in the following layer, added context to what was observed in layer one in form of talk (Peräkylä 2005). As the 'professional vision' (Goodwin 1994) of this paper was to privilege other dimensions than talk, heuristics (Norris 2012) were chosen as a

method for conveying dimensions such as the pitch employed and how speech coincided with movement. This added directionality (Bezemer & Mavers 2011) and depth to the emotional and qualitative descriptions of movement from layer one, as opposed to other methods of transcription (e.g. Heath et al., 2010a, b) which lets the reader retain more agency by letting them design a course of their own in reading the transcripts.

The properties of and objects in the hallway, such as the length of the hallway with its markings and the pictures on the wall are not neutral, and function as mediators that configure (Roehl 2012) and order (Geertz 1973) interaction. As such, the environment with its materiality is a central component to consider when wanting to understand human interaction. Yet reflecting on the analysis of the environment in this paper, I am reminded of Schegloff's (1991) 'criterion of relevance' which states that what is relevant must be shown to be relevant to the actors. The issue with noting the lines of the floor or the pictures on the wall is that although they have been significant in shaping the researcher's ethnographic knowledge of this place, the students do not draw attention to these in their visible orientation. On the one hand, these artefacts act as a resource to understand how the idea of running as fast is introduced into the task, but on the other hand, we can at the same time question if this is not an over-interpretation on the part of the researcher. This is where the voices of the participants become crucial. They are central to understanding the environment and avoid over-interpretation of interaction (Knoblauch 2009). In the above analysis this meant read-justing the understanding of Alfons act of sitting down, from an act born primarily out of his unwillingness to run to also take into considerations his habits when situated in the hallway. While the importance of emic perspectives for video analysis is already highlighted (Knoblauch 2009), pulling it in as an analytical layer is new and adds transparency to growing interpretation.

While there seems to be a strong focus on structured and systematic approaches to working with video in general (Derry 2007; Erickson 2006; Knoblauch et al. 2015) and video observations in and across classrooms in particular (Klette 2009), there seems to be a paucity of structure when it comes to the process of interpreting and analyzing particular segments holistically. Even in state-of-the-art methodologies (Knoblauch 2012; Knoblauch et al. 2015), the process of transcription remains elusive. Knoblauch et al. (2015) broadly refer to the process as *detailed transcription* and from looking at their transcripts, it is clear that different layers are foregrounded and merged in horizontal diagrams. Yet, it remains unclear how each element in the diagram came into being and how these elements shape the interpretations made. The methodology in this paper wrestles with this issue, as it attempts to create clarity and structure to the process of inferring on the background of different modes of data. Albeit, by addressing one layer at the time it is possible to show how interpretation of an event grow and are informed by the insights from each layer.

Developed within an interpretative and qualitative framework for working with visual analysis, this approach is suited for naturalistic footage of human interaction where the researcher seeks to come closer an understanding of what is taking place and how the participants draw on embodied and situated resources when acting.

Conclusion

The aim of this article was to propose an analytical framework for working systematically with video, emphasizing in particular embodied dimensions of video data. By taking into consideration how the body and its emergence can never be fully captured or understood by a single theoretical perspective, as different theories foreground different qualities of the body, the notion of *layering* was conceived as a way forward to merge different perspectives into more holistic understandings of embodied interaction. *Layering* as a methodology for analyzing video was presented using the metaphor of *the onion*. Inspired by the idea of a interaction as aggregates of layers, the approach advocated for a process of systematically peeling (identifying layers) and merging layers (analyzing and combining in aggregates) to qualify and substantiate interpretation and understanding when working with video. As such, this method sought to bring transparency and structure into qualitative research. Based on the exemplification of the approach with data obtained from a classroom-based study, I claim that different layers can be combined successfully in merged transcripts and that this structured approach provides a holistic impression of the embodied dimensions in the video segment. This is in part because privileging visual data as the first layer enables a focus on more affective features of interaction, where the body is not reduced to a single feature of the body. Yet in part also due to the strong focus to merge the layers in transcripts, which prompted the identification of ways of representing and communicating talk, environment and emic perspectives that added to and deepened the previous layers. Despite an inherent focus on transparency when making interpretations in the layered approach itself, a central insight gained when working with the method was the value of emic voices as a way to support and/or readjust interpretations. This article considered a very particular order for taking note of embodied dimensions, but it did not consider the consequence for analysis of trying out a different order. Future research may therefore wish to explore a different order of those layers to consider how a different order or additional layers may affect the analysis and interpretation of a given segment.

Endnotes

[1]The Doppler Effect is the change in frequency of a wave for an observer who is moving relative to its source. This is a phenomenon an observer will experience when a siren passes him, where the sound will transform from a higher pitch during the approach, true at the moment it passes by, and lower when it recedes (for more information see e.g. www.grc.nasa.gov/www/k-12/airplane/doppler.html)

[2]Kinesphere is the 3D volume of space that I can access with my body without shifting my weight to change my stance (Laban & Lawrence, 1974)

Authors' contributions

The author read and approved the final manuscript.

Authors' information
Liv Kondrup Kristensen is an adjunct assistant professor at the University College Absalon and Ph.D. student at Aalborg University, Denmark.

Ethics approval and consent to participate
The author confirms that this research project has followed international best practices on ethical conduct in educational research, in particular following the Danish Code of Research Integrity. All participants involved have been provided with a letter of informed consent and all those whose data has been shared in this publication have consented through their signature on the above stated form to sharing the information concerning them.

Competing interests
The author declares that she has no competing interests.

Publisher's Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References
Ayass R (2015) Doing data: the status of transcripts in conversation analysis. Discourse Studies 17(5):505–528 https://doi.org/10.1177/1461445615590717
Bezemer J, Cope A, Kress G, Kneebone R (2014) Holding the scalpel: achieving surgical Care in a Learning Environment. J Contemp Ethnogr 43(1):38–63 https://doi.org/10.1177/0891241613485905
Bezemer J, Mavers D (2011) Multimodal transcription as academic practice: a social semiotic perspective. Int J Soc Res Methodol 14(3):191–206 https://doi.org/10.1080/13645579.2011.563616
Clarke, D., Keitel, C., & Shimizu, Y. (2006). Mathematics classrooms in twelve countries: The insider's perspective (Vol. 1.). Sense Publishers
Cohen L, Manion L, Morrison K (2011) Research methods in education, 7th edn. Routledge, London
Cowie, B., Otrel-Cass, K., & Moreland, J. (2012). Finding Out about Fossils in an Early Years Classroom: A Context for Developing a "Practical Explanatory Theory." In B. Kaur (Ed.), Understanding teaching and learning (pp. 159–169). Rotterdam: Sense Publishers
Derry, S. J. (2007). Guidelines for conducting video research in education: Recommendations from an expert panel. Retrieved January 31, 2018, from https://drdc.uchicago.edu/what/video-research-guidelines.pdf
Derry, S. J., Pea, R. D., Barron, B., Engle, R. A., Erickson, F., Goldman, R., … Sherin, B. L. (2010). Conducting video research in the learning sciences: guidance on selection, analysis, technology, and ethics. Journal of the Learning Sciences, 19(1), 3–53. https://doi.org/10.1080/10508400903452884
Dreyfuss H (1991) Being-in-the-world. A commentmy on Heidegger's "being and time". MIT Press, Cambridge, MA
Erickson F (2006) Definition and analysis of data from videotape: Some research procedures and their rationales. In: Green JL, Camilli G, Elmore PB, Skukauskaitė A, & Grace E (eds) Handbook of Complementary Methods in Education Research. Routledge, New Jersey, p 177–192
Geertz C (1973) The interpretation of cultures: selected essays. Basic Books, New York
Glaser BG, Strauss AL (1967) The discovery of grounded theory. Aldine, Chicago
Goffman E (1959) The presentation of self in everyday life. Penguin Books, London
Goodwin C (1981) Conversational organization: interaction between speakers and hearers. Academic Press, New York
Goodwin C (1994) Professional vision. Am Anthropol 96(3):606–633 https://doi.org/10.1525/aa.1994.96.3.02a00100
Goodwin C (2000) Action and embodiment within situated human interaction. J Pragmat 32(10):1489–1522 https://doi.org/10.1016/S0378-2166(99)00096-X
Goodwin C (2007) Participation, stance and affect in the organization of activities. Discourse & Society 18(1):53–73 https://doi.org/10.1177/0957926507069457
Goodwin C (2009) Embodied hearers and speakers constructing talk and action in interaction. Cognitive Studies 16(1):51–64 https://doi.org/10.11225/jcss.16.51
Goodwin C, Goodwin MH (2004) Participation. In: Duranti A (ed) A companion to linguistic anthropology. Blackwell Publishing, Oxford, pp 222–245
Goodwin MH, Cekaite A, Goodwin C (2012) Emotion as stance. In: Sorjonen M-L, Peräkylä A (eds) Emotion in interaction. Oxford University Press, Oxford
Heath C, Hindmarsh J (2002) Analysing interaction: video, ethnography and situated conduct. In: May T (ed) Qualitative research in action. Sage, London, pp 99–121
Heath C, Hindmarsh J, Luff P (2010a) Video in qualitative research. Sage, London
Heath C, Hindmarsh J, Luff P (2010b) Video in qualitative research. Analysing social interaction in everyday life. SAGE, Los Angeles, Calif
Heath C, Luff P (2000) Technology in action. Cambridge University Press, Cambridge
Hutchby I (2003) Affordances and the analysis of technologically mediated interaction: a response to Brian Rappert. Sociology 37(3):581–590 https://doi.org/10.1177/00380385030373011
Jacobs JK, Kawanaka T, Stigler JW (1999) Integrating qualitative and quantitative approaches to the analysis of video data on classroom teaching. Int J Educ Res 31(8):717–724 https://doi.org/10.1016/S0883-0355(99)00036-1
Jefferson G (1984) Transcription notation. In: Atkinson J, Heritage J (eds) Structures of social interaction. Cambridge University Press, New York
Kendon A (2004) Gesture: visible action as utterance. Cambridge University Press, Cambridge

Klette K (2009) Challenges in strategies for complexity reduction in video studies. Experiences from the PISA+ study: A video study of teaching and learning in Norway. In: Janik T & Seidel T (eds) The Power of Video Studies in Investigating Teaching and Learning in the Classroom. Waxmann, Münster, p 61–82

Knoblauch, H. (2009). Social construtivism and the three levels of video analysis. In U. T. Kissmann (Ed.), Video Interaction Analysis: Methods and Methodology (pp. 181–198). Frankfurt am Main: Peter Lang

Knoblauch, H. (2012). Videography: Focused Ethnography and Video Analysis. In H. Knoblauch, B. Schnettler, J. Raab, & H.-G. Soeffner (Eds.), *Video analysis, methodology and methods: qualitative audiovisual data analysis in sociology* (3rd ed., pp. 69–83). Frankfurt am Main: Peter Lang

Knoblauch, H., Schnettler, B., & Raab, J. (2012b). Video-Analysis. Methodological Aspects of Interpretive Audiovisual Analysis in Social Research. In H. Knoblauch, B. Schnettler, J. Raab, & H.-G. Soeffner (Eds.), *Video analysis: Methodology and Methods. Qualitative Audiovisual Data Analysis in Sociology* (3rd ed., pp. 9–27). Frankfurt am Main: Peter Lang

Knoblauch H, Schnettler B, Raab J, Soeffner H-G (eds) (2012a) Video analysis: methodology and methods, 3rd edn. Frankfurt am Main, Peter Lang

Knoblauch H, Tuma R, Schnettler B (2015) Videography: introduction to interpretive videoanalysis of social situations (Elektronis). In: Frankfurt am Main. Peter Lang, Germany

Konie, R. (2011). A brief overview of Laban Movement Analysis. Retrieved July 5, 2017, from http://www.movementhasmeaning.com/

Kvale, S., & Brinkmann, S. (2008). Interview: introduktion til et håndværk (2nd ed.). Kbh.: Hans Reitzel

Laban, R., & Lawrence, F. C. (1974). Effort: economy of human movement (2. ed.). Plymouth: Macdonald and Evans

Laban, R., & Ullmann, L. (1975). The mastery of movement (3. ed.). Boston: Plays

Laurier, E., & Philo, C. (2012). Natural Problems of Naturalistic Video Data. In H. Knoblauch, B. Schnettler, J. Raab, & H.-G. Soeffner (Eds.), Video analysis: Methodology and Methods (3rd ed., pp. 181–190). Frankfurt am Main: Peter Lang

Luckmann, T. (2012). Some Remarks on Scores in Multimodal Sequential Analysis. In H. Knoblauch, B. Schnettler, J. Raab, & H.-G. Soeffner (Eds.), Video analysis, methodology and methods: qualitative audiovisual data analysis in sociology (3rd ed., pp. 29–34). Frankfurt am Main: Peter Lang

Merleau-Ponty M (2012) Phenomenology of perception, trans.C. Smith. Routledge, London

Morgan A (2007) Using video-stimulated recall to understand young children's perceptions of learning in classroom settings. European Early Childhood Education Research Journal 15(2):213–226 https://doi.org/10.1080/13502930701320933

Norris S (2004) Analyzing multimodal interaction: a methodological framework. Routledge

Norris S (2012) Teaching touch/response-feel. A first step to an analysis of touch from an (inter)active perspective. In: Norris S (ed) Multimodality in practice: investigating theory-in-practice-through-methodology. Routledge, New York, pp 7–19

O'Loughlin M (1998) Paying attention to bodies in education: theoretical resources and practical suggestions. Educational Philosophy and Theory 30(3):275–297 https://doi.org/10.1111/j.1469-5812.1998.tb00328.x

Peräkylä, A. (2005). Analyzing talk and text. In N. K. Denzin & Y. S. Lincoln (Eds.), The SAGE handbook of qualitative research (3. ed., pp. 869–886). Thousand Oaks, Calif.: Sage Publications

Polanyi M (2009) The tacit dimension. University of Chicago Press, Chicago

Raab, J., & Tänzler, D. (2012). Video Hermeneutics. In H. Knoblauch, B. Schnettler, J. Raab, & H.-G. Soeffner (Eds.), *Video analysis: Methodology and Methods* (3rd ed., pp. 85–97). Frankfurt am Main: Peter Lang

Radinsky J (2007) Student-created artifacts: mediators of science discourse and reasoning with visual data. In: Paper presented at the annual meeting of the American Educational Research Association (AERA). April, Chicago, IL

Raingruber B (2003) Video-cued narrative reflection: a research approach for articulating tacit, relational, and embodied understandings. Qual Health Res 13(8):1155–1169 https://doi.org/10.1177/1049732303253664

Ritchie SM, Newlands JB (2016) Emotional events in learning science. In: Bellocchi A, Quigley C, Otrel-Cass K (eds) Exploring emotions, aesthetics and wellbeing in science education research. Springer, Basel, pp 107–119

Roehl T (2012) From witnessing to recording – material objects and the epistemic configuration of science classes. Pedagogy, Culture & Society 20(1):49–70 https://doi.org/10.1080/14681366.2012.649415

Schegloff EA (1991) Reflection on talk and social structure. In: Boden D, Zimmerman D (eds) Talk and social structure. Cambridge, Polity, p 44–70

Schmitt R, Deppermann A (2007) Monitoring und Koordination als Voraussetzungen der multimodalen Konstitution von Interaktionsraumen. In: Schmitt R (ed) Koordination. Studien zur multimodalen Interaktion. Tilbingen, Gunter Narr Verlag, pp 95–128

Schnettler, B. (2012). Orchestrating Bullet Lists and Commentaries. A Video Performance Analysis of Computer Supported Presentations. In H. Knoblauch, B. Schnettler, J. Raab, & H.-G. Soeffner (Eds.), *Video analysis: Methodology and Methods* (3rd ed., pp. 155–169). Frankfurt am Main: Peter Lang

Schnettler, B., & Raab, J. (2008). Interpretative Visual Analysis. Developments, State of the Art and Pending Problems. *Forum: Qualitative Social Research*, *9*(3), Art. 31

Soeffner, H.-G. (2012). Visual Sociology on the Basis of "Visual Concentration." In H. Knoblauch, B. Schnettler, J. Raab, & H.-G. Soeffner (Eds.), *Video Analysis: Methodology and Methods* (3rd ed., pp. 205–217). Frankfurt am Main: Peter Lang

Streeck J (2003) The body taken for granted: lingering dualism in research on social interaction. In: Glerm PJ, LeBaron CD, Mandelbaum J (eds) Studies in language and social interaction. Erlbaum, Mahwah, NJ, pp 427–440

Streeck J (ed) (2014) Embodied interaction (learning in doing: social, cognitive and computational perspectives) - language and body in the material world. Cambridge University Press, Cambridge

Streeck J, Goodwin C, LeBaron C (2011) Embodied interaction in the material world: an introduction. In: Streeck J, Goodwin C, Lebaron C (eds) Embodied interaction. Language and the body in the material world. Cambridge University Press, Cambridge, UK, pp 1–26

Suchman LA (1987) Plans and situated actions: the problem of human machine communication. Cambridge University Press, Cambridge

Thøgersen U (2014) The embodied emotionality of everyday work life: Merleau-Ponty and the emotional atmosphere of our existence. Philosophy of Management 13(2):19–31 https://doi.org/10.5840/pom20141329

Zembylas M (2007) The "specters" of bodies and affects in the classroom: a Rhizo-ethological approach. Pedagogy, Culture and Society 15(1):19–35 https://doi.org/10.1080/14681360601162030