Qualitative Analysis of Video Data: Standards and Heuristics

Kay E. Ramey (chair), Northwestern University, kayramey@u.northwestern.edu
Dionne N. Champion, Northwestern University, dionnechampion2012@u.northwestern.edu
Elizabeth B. Dyer, Northwestern University, elizabethdyer@u.northwestern.edu
Danielle T. Keifert, Exploratorium, dkeifert@exploratorium.edu
Christina Krist, Northwestern University, ckrist@u.northwestern.edu
Peter Meyerhoff, Northwestern University, peter.meyerhoff@u.northwestern.edu
Krystal Villanosa, Northwestern University, kvillanosa@u.northwestern.edu

Jaakko Hilppö (discussant), Northwestern University, jaakko.hilppo@northwestern.edu

Abstract: Video research is an increasingly important method in the learning sciences. Video provides unique analytical affordances to researchers but also presents unique tensions, many of which have not yet been adequately addressed in the literature. The authors of this symposium draw on their diverse experiences, analyzing a variety of video corpuses, to provide theoretical and methodological standards and heuristics for the process of video analysis. We focus on three themes central to the process of video analysis that would benefit from increased theoretical and methodological attention: transcription tensions, defining the unit of analysis, and representing context. We discuss how our approaches to video analysis are framed by theory and how we have applied them to specific datasets, to answer a variety of research questions. In doing so, we make explicit some crosscutting methodological norms and invite continued discussion about these norms from multiple analytic traditions.

Keywords: video, qualitative, methods, analysis, data representation

Session summary

Researchers increasingly rely on video records to analyze the processes of teaching and learning. Video provides both breadth (footage that spans weeks or months of activity) and depth (a richly detailed, moment-to-moment interactional record). Video creates powerful new affordances, such as the ability to rewind or to see multiple participant perspectives concurrently, that traditional qualitative research methods (e.g., Glaser & Strauss, 1967; Miles & Huberman, 1994) generally do not afford. The literature on using video in the study of teaching and learning provides guidance for researchers in selecting, capturing, and representing video data (see Derry, et al., 2010, for a recent review). In particular, there are rich commentaries on approaches to progressively refining hypotheses (Engle, Conant & Greeno, 2007), conceptualizing the epistemology of video representations (Goldman, 2007), and representing video data in ethnographically adequate ways (e.g., McDermott, Gospodinoff, & Aron, 1978; Ochs, 1979). However, the field lacks consensus on theoretical and practical guidelines for the *process* of video analysis: asking the difficult, multilayered questions of "what do I do with all of this video?" and "how do I go about doing it?" In this symposium, we integrate guidelines from specific analytic traditions (e.g., interaction analysis) and general methodological approaches to qualitative data analysis (e.g., Miles, Huberman & Saldana, 2014). We discuss the reflexive relationship between various learning theories, existing methodologies, and the methods that we have developed to address the unique affordances and challenges of analyzing video data.

Each of the authors draws on experience working with a different corpus of video data. Our research interests are wide-ranging and include middle-school science instruction, museum interactions, at-home family inquiry, teacher practices, spatial thinking, design of learning spaces, and learning through dance. Across these different data sets and research questions, however, we have recognized a common set of core challenges in video data analysis that are not adequately addressed by current literature on video methods. Through workshop collaborations, we have identified issues of practical and theoretical concern for our video research, and we have developed a set of standards and heuristics to work through these challenges that cut across different research projects and analytic traditions. Our findings, and this symposium, are organized into three areas of focus:

1. *Transcription tensions*. Which interactional phenomena need to be transcribed? When? How? We consider these questions in the context of our work, characterizing transcription as an analytic and interpretive act, central to ongoing analysis, and propose a recursive method of video transcription.

We also show how transcription can be used as a tool in the analytic process and demonstrate a technique for foregrounding visual phenomena during analysis.

2. *Defining the unit of analysis*. The richness of video data often results in multiple rounds of analysis that include revisions to theoretical frameworks. It can also lead to separate lines of analysis using distinct theoretical frameworks. Even given the same video data, changes in frameworks can lead to changes in codes AND in the units being coded (e.g., individual utterances; episodes of talk + gesture). We discuss strategies for maintaining flexibility and issues in reporting reliability with these kinds of analyses.
3. *Representing context*. Video data afford unique access to context surrounding phenomena of interest, but there is little guidance on how to incorporate this context into analysis. We explore different ways that context can be represented and analyzed using video data, as constituted both by the researcher and by participants.

These three areas of focus are not novel; in fact, they are some of the first issues with which novice qualitative researchers will grapple. However, video data provide new possibilities for how to address each issue. Although norms for utilizing these new possibilities exist in practice within various analytic sub-communities, they are not explicit or available to the broader research community except as individual case studies exemplifying particular approaches. For example, Engle, Conant, & Greeno (2007) discuss the general strategy of progressively refining hypotheses. They also provide a specific example of how they transcribe, select units of analysis, and represent context. However, that example does not provide general guidance for *how* to make or evaluate those decisions throughout the analytic process. Our goal in this symposium is to make these decision-making norms explicit as general strategies and heuristics. Much like how proposing theoretical mechanisms for scientific phenomena then allows for empirical investigation of those phenomena (Machamer, Darden, & Craver, 2000), proposing theoretically-grounded strategies and heuristics for video analysis allows for refinement, revision, or rejection of these strategies. We view this as a critical process for the field for continued development of methodological rigor around analysis of video data.

We will structure this symposium as an interactive session. We will begin with a 5-minute introduction to the affordances, tensions, and open questions in video analysis. This will be followed by 10-minute presentations by the authors of each sub-section, covering their theoretical and methodological standards and heuristics by situating them within concrete examples from their data corpuses. Following the presentations, the discussant will present questions for discussion. These will guide interaction at distributed stations during the next 50 minutes, where the authors will present more detail about how they have applied these standards and heuristics to their specific analytic contexts and invite feedback and discussion from session participants. We will conclude with brief summary remarks from the discussant and time for whole-group discussion.

Transcription tensions

Danielle Keifert, Exploratorium; Kay E. Ramey, Peter Meyerhoff, and Christina Krist, Northwestern University

Video uncovers a wide range of interactional modalities; people use talk, gesture, gaze, body position, facial expression, movement, and material objects to exchange ideas and information (Goodwin, 2013; Hall, 1999). How can the complexity and dynamism of knowledge in use (Hall & Stevens, 2016) be captured in static representations, and how does creating those representations shape our understanding of interactions? Researchers have long understood that depictions of action through transcription are theoretical in nature (Ochs, 1979); transcription decisions can be political (Bucholtz, 2000) and position research within a particular tradition (Bezemer & Mavers, 2011). To create a transcript is to make consequential choices about which phenomena merit representation. This relies upon our understanding of what is happening in the interactions we represent. The decision to transcribe in a particular style cues a specific lens on what happens in activity; transcribing words exclusively, for example, obscures all other interactional phenomena. The transcript is not a neutral piece of objective data but rather the product of an analytic move, in which the researcher selects one or more interactional modalities to focus on in the analytic process (Bezemer & Mavers, 2011). Goodwin (2003) recognized "recursive interplay between analysis and methods of description" (p. 161), as the researcher views, re-views, and documents video-recorded activity through multiple lenses, to progressively explore and develop an argument. Different, equally valid, transcripts can be produced from the same video record, reflecting differences in research questions, analytic frames, and phenomena of interest. Transcripts evolve as researchers work to develop arguments, and the process of working through multiple lenses can support more systematic analysis of the multiple modalities of knowledge use. Creating multiple transcripts during the analytical *process*, regardless of the final transcript *product*, supports researchers' developing understandings of interaction. Here, we build on Goodwin's idea of

recursion to demonstrate a method of *recursive transcription,* a theoretically-grounded heuristic for exploring knowledge in use.

Recursive transcription

Video offers an open invitation to the researcher to look beyond the spoken word and find meaning from other dimensions of participant activity. Researchers recommend being sensitive to and intentional about which semiotic fields are chosen for transcription (Bucholtz, 2000; Derry et al., 2010; Goodwin, 2013), though researchers mostly focus on such representations in final products (e.g., articles, presentations). We propose that the development of detailed transcripts representing multiple semiotic fields is more than a matter of representational choice in final product: it can guide the process of noticing during analysis. By engaging in *systematic,* sequential analysis of verbal and nonverbal interactional phenomena, the researcher must consider the possibility of meaning in modalities other than talk. Drawing on Stevens (2012), Hall (1999), and Goodwin (2013), we identify these nonverbal, semiotic fields of interest as: gesture and pointing, gaze and attention, body position and movement, touch, tone and inflection, facial expression, and engagement with material objects.

For example, in a study by Keifert (2015), she selected a moment for close analysis in which toddler Catherine asked Dad about a thermometer. In a first transcription pass, she documented content of talk as "Dad: See the red in there. The red line? If the red is, if the red line goes up here in this red area it's hot." She then re-viewed the video repeatedly, attending on each pass to a different modality represented above. This "building-up" process produced a detailed multimodal transcript in narrative form (see partial transcript below). Critically, this process encouraged Keifert to notice interactional phenomena other than talk. For instance, in documenting that Catherine turned away as her brother splashed nearby, then quickly turned back to look at Dad's gesture (see below "[6]here in this red area"), the researcher was positioned to notice Catherine's sustained interest (looking again after her brother quit splashing) although Catherine did not express interest verbally. Through considering and transcribing a broad range of modalities in her analysis, Keifert determined that the following indications of talk, gaze, touch, gesture, and engagement with objects merited inclusion in the final transcript:

> Dad: [1]See the [2]red in there. [3]The red line? [4]If the red is, [5]if the red line goes up in [6]here in this red area it's hot.
>
> [1] *Catherine looks at the thermometer as Dad touches the thermometer.*
> [2] *Dad uses his finger to point to the red line in the middle of the thermometer.*
> [3] *Dad runs his finger up the middle of the thermometer from top to bottom.*
> [4] *Catherine turns her head away from her brother splashing.*
> [5] *Splashing stops and Catherine turns back towards Dad and the thermometer.*
> [6] *Dad points and makes a small circle pointing to the top of the thermometer.*

An alternate move would have been to focus on the thermometer as a scientific artifact, for example representing the number scale to which the participants oriented their understanding of temperature.

Having decided *what* to transcribe, the decision turns to *how*. Jefferson's (2004) conventions for transcribing talk are widely accepted in the field of conversation analysis, but a range of techniques have been proposed to address the challenges of multimodal transcription (Jewitt, 2009; Bezemer & Mavers, 2011). During the transcription process, it may be sufficient to translate nonverbal activities into narrative verbal descriptions (as above). However, for some interactional data and analytic frames, an alternative approach is needed, which circumvents the translation errors inherent in re-representing visuospatial modalities (e.g., gesture) in words. We refer to this approach as *visual transcription.*

Visual transcription

Figure 1 illustrates three ways of representing the moment Dad oriented Catherine to the red line in a visual transcript. When representing interactions, communicating action and spatial relations is a key challenge. Each of these representations foregrounds a different aspect of activity: the still provides a detailed visual inventory of people, objects, and space within the camera's frame; the sketch focuses on body positioning and action (e.g., removing the squirt gun in Dad's left hand, adding arrows indicating Dad's motion); the verbal transcript reports spoken words. Keifert created multiple visual transcripts, in addition to these, to explore this one moment during analysis, but ultimately decided on a combination of 1B and 1C to support the claim that Dad made "hot" and "cold" relevant extremes when talking about temperature. This process of recursive visual transcription supported her identification of all elements of the interaction—verbal (e.g., Dad's words) and nonverbal (e.g., Catherine, Dad, thermometer, gaze, body positions, and gestures)—allowing her to identify and edit out distractors (e.g.,

gun, pool, brother). Such streamlining allowed her, and her audience to focus on features of the interaction made relevant by participants and related to her argument. Thus, visual transcripts foreground analytical decisions that might otherwise be left implicit, and serve as documentation of those decisions.

1A

1B

Dad: If the red is, if the red line…

1C

Figure 1. Representations of video data as a video still (1A), a sketch (1B), and a verbal transcript (1C).

This example highlights a particular limitation of video transcription: the challenge of representing gesture. Whereas researchers working with talk can draw on a rich representational scheme from linguistics and conversation analysis (Jefferson, 2004), there are few such *explicit* standards for depicting gesture. We encourage further work to develop explicit transcription techniques—*in process,* to guide analysis, and *in product,* to guide readers—to capture gesture as systematically as verbal communication.

Lessons learned

Though the above examples focused on one moment of interaction, the authors here study diverse phenomena in diverse contexts. Each of us has approached the complexity of transcription in our own way, but the process of recursive, visual transcription has led us all to new insights similar to the examples above. We believe that our experiences with transcriptions of multiple semiotic fields provide the following insights for the field: (1) that the "what, when, and how" of transcription can and should remain in flux as we develop our arguments, because (2) it is precisely through engaging with these decisions that we come to understand our data more clearly. In a field lacking in both consensus and detailed, process-based guidelines for transcribing and analyzing video data, we encourage recursive, visual transcription as an integrative standard for diverse methodological approaches.

Defining the unit of analysis

Christina Krist, Krystal Villanosa, and Kay E. Ramey, Northwestern University

Scholars using video data maintain that how researchers approach video analysis depends on their theoretical commitments and the specific research questions they are pursuing (Barron & Engle, 2007; Goldman, Erickson, Lemke, & Derry, 2007). We add that the richness and complexity of video data often results in iterative cycles of analysis that include revisions to theoretical frameworks. It can also lead to concurrent or asynchronous analyses that employ distinct theoretical frameworks. In both cases, changes in frameworks lead to changes in how we define our unit(s) of analysis—meaning we make changes not just to our codes, but to *what* we code. These changes have deep implications for how we determine reliability and think about representativeness. Here, we discuss how theoretical frameworks determine what phenomena are analyzed, how they are analyzed, and how we might re-think reliability. We place particular emphasis on the fluidity of unit(s) during iterative cycles of analysis, highlighting both the practical and theoretical challenges we encounter in our work.

To ground our discussion of how theoretical frameworks shape our interpretation of video data, particularly how we define our unit(s) of analysis, we first detail data collected from a study conducted in a museum. These data consist of two sets of video recordings. The first set is of parent-child dyads playing an interactive game. The second is of parent-child dyads discussing objects in two exhibit cases. These data were collected to understand the effect of games on museum visitors' conversations around exhibited objects. The two theoretical frameworks we applied to these data were conversational elaboration (CE; Leinhardt & Crowley, 1998), and cultural forms (Saxe, 1999). CE is defined as visitor talk taking place during a museum visit and focuses on how meaning, experiences, and interpretations develop. Cultural forms originates from Saxe's form-function shift framework. It describes how social constructions, conventions, and systems of representation (e.g., currency, games) are restructured through social participation, taking on novel functions.

Although applied to the same corpus of video data, these two frameworks result in vastly different units of analysis. CE leads us to code for moments when parent-child dyads' conversations become "elaborate" by an

expansion of details. As our analysis progressed and we re-watched videos, we learned that we needed to refine our codes so that our units were coupled with the particular exhibited objects that parent-child dyads were visually attending to, as different objects afford different kinds of elaborations. In this instance, while our theoretical framework and research questions governed what we coded for initially, we used feedback from our video recordings to iterate on our codes and refine our units of analysis. Even within the same framework, units of analysis can change because of what video data reveal about the phenomena under investigation.

In the process of conducting our analysis using the CE framework, the wide range of social interactions and the concrete details of the physical artifacts present in our video recordings allowed us to conduct a secondary analysis, ask new questions of these data, and apply a new theoretical framework—cultural forms. In particular, we became interested in how parent-child dyads were adopting and potentially restructuring different design elements in the game to advance through its levels. Consequently, we coded how parent-child dyads' used game pieces, interpreted rules, and applied strategies to maneuver through the game. Using cultural forms as an analytic lens led us to code a different set of episodes than we were coding under the CE framework.

Given the multiple possibilities for units of analysis in video data, reliability and representation become increasingly complicated. As a field, we struggle with making analytical decisions explicit in a way that communicates reliability and representativeness of our claims. Here, we present a definition of reliability that includes explicit criteria for "fuzzy" reliability, to help us move towards increased rigor in analytic claims.

To ground this discussion, we used analysis of video data collected from science classrooms, as a part of a study examining how students learn to engage in scientific practices in meaningful ways. As in the work described above, we found decisions about what counts as a unit to be as consequential as decisions about which code(s) should be applied to the unit. As a result, we spent a significant amount of time coding for, discussing, and checking for agreement on the identification of units themselves. These checks for "fuzzy" reliability relied on heuristics for determining whether we were seeing the same things in the data. Figure 2 illustrates what this process involved. In this case, we used three checks for "fuzzy" reliability early on in the coding process, to ensure that we agreed on the same instances for further analysis. Once we agreed upon those instances, we coded them and used traditional calculations for reliability. In addition, the "fuzzy" checks provided leverage, in thinking about and communicating representativeness. They served as a trace of our work through the data that allowed us to not only talk about how representative a particular code was within the set of coded episodes, but also to discuss in a multi-tiered way how representative those episodes were within the dataset.

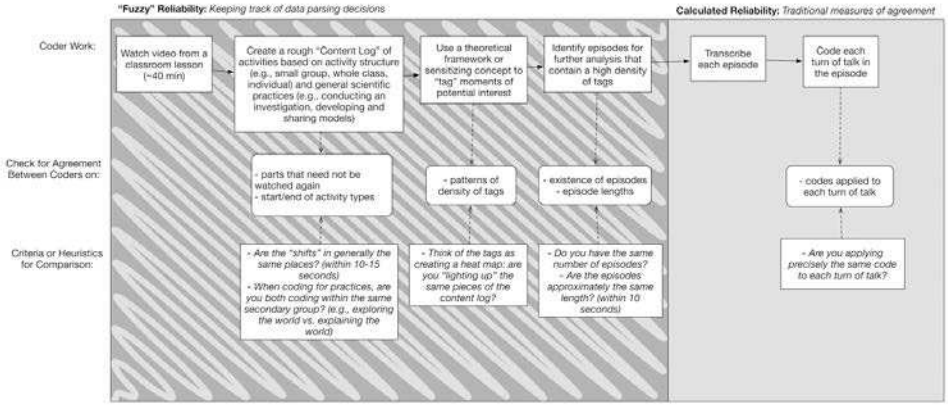

Figure 2. Sample heuristics for comparison of coding decisions in the video coding process.

Lessons learned

Video data provide a range of possibilities for units of analysis and give researchers opportunities to conduct rich secondary analyses. The work described here highlights how the operationalization of theoretical frameworks during iterative processes of analysis influences the ways in which researchers develop and refine units of analysis. The connection between analytic frameworks and units of analysis should be made explicit when presenting data, as it greatly influences the nature of our findings. However, acknowledging different units for different analyses provides additional challenges with regards to reliability. We have drawn attention to a second version of reliability, a "fuzzy" reliability, that involves tracking and checking for agreement on multiple analytic decisions about units of analysis before relying on traditional statistical calculations of reliability .

Representing context

Dionne N. Champion and Elizabeth B. Dyer, Northwestern University

A broad range of work in the field (Vygotsky, 1978; Scribner, 1984; Lave 1988; Hutchins, 1995) has shown that to fully understand how people learn, it is important to look beyond the individual, pointing to the importance of understanding interactions and relationships between individuals, artifacts, and social groups. Because video is a comprehensive data source—simultaneously capturing talk, gaze, gesture, movement, and interactions in a format that is available for repeated viewing—it can be an ideal source for capturing and analyzing context. However, video data can easily become overwhelming to the researcher who must sift through hours of video, deciding which details of the context are significant. In this paper, we introduce strategies for grappling with these issues as we consider context in the analysis of learning events in situ. We draw upon ideas from Actor Network Theory (Latour, 2005), a social context theory which suggests that researchers interrogate their data without presupposing an understanding of which factors matter, and instead look closely at what is happening and trace the observable associations. This stance is essential to understanding the whole story told by the data.

In order to provide a context for this discussion, we share examples from video data collected during an informal, in-school STEM learning program for 5th and 6th grade students. The setting has a complex activity structure, in which all participants engaged in different activities in a variety of ways. The activities were emergent, interest-based, and child-driven, with little teacher direction. The semi-structured activities gave students freedom of choice for decision-making and interaction. Video data in this context was collected both by a whole room, stationary camera and by focal students wearing visor cameras.

Accounting for multiple perspectives

Research on collecting video data has established that video is not a "neutral" source of data; inherent in the collection of video is the perspective of the camera and the researcher's decisions about which perspective to foreground in the data (Derry et al., 2010; Goldman, 2007). Although the collection of video inherently influences the perspective of the data collected, video can still allow for flexible perspective-taking during analysis. In our process of analysis, clips are selected based on research questions or on emerging themes discovered in the data. Methods of interaction analysis (Hall & Stevens, 2016) are used to understand what is consequential to the task or to a participant's goal in a moment of interest, and micro-analysis leads to choices about what is important to trace. From there, the view is expanded by looking at other related clips and/or camera angles. Some examples of expanding the view include: tracing what happens just before or after the moment of interest; close analysis of other participant views (e.g. from other participants wearing visor cameras) to fill in details about interactions with others; and analysis of wider angle perspective cameras to trace participant pathways and association networks.

Our example, partially transcribed below, demonstrates the value of considering multiple perspectives. It involves two boys wearing visor cameras, one who seeks out the other for help.

1	Akeem: Okay- Oh no, oh no! I'm gonna walk.	(*Walks across the room to Benji and Evan)*
2	Akeem: Hey we're camera bros! What's up	(*Reaches across Evan toward Benji's face, to high five Benji)*
3		(*Benji looks up awkwardly at Akeem)*
4	Akeem: Uh can you help me with something?	
5	Benji: We're working...	
6	Akeem: Okay I'm gonna ask Brian.	(*Walks away from Benji and Evan)*

This first transcript excerpt is taken from the point-of-view camera worn by Akeem. A close, multimodal analysis of this perspective allowed the researcher to attend to Akeem's talk, to his focus as he works, and to his level of engagement in the activity. It also allowed the researcher to see both a first-hand perspective of the materials and other actors integral to his process and the moment when his request for help is shut down (line 5). If focusing only on the initial perspective, the researcher could conclude that Benji had no interest in helping Akeem. In order to incorporate the context surrounding Akeem's request for assistance, it is helpful to analyze the actions of multiple participants, as well as the more complete picture of what is happening in the room. Widening the lens by analyzing data from multiple cameras helped to paint a fuller picture of this event.

Benji's Perspective	Akeem's Perspective
Benji: We got it! So where do we tape the thread?	Akeem: I'm gonna walk (*walks across the room*)

Evan: Okay, so...	
Benji: Feel free to flip the light and tape the thread to-	Akeem: Hey we're camera bros! What's up (r*eaches across to high five Benji)*
Benji: *(looks up awkwardly at Akeem)*	Akeem: Uh can you help me with something?
Benji: We're working together right now	

Including Benji's perspective, as captured by his visor camera, in the analysis helped clarify his response and brings some context to his decision to say no. The added perspective helped the researcher understand that Benji was highly engaged in his own process at the time of Akeem's question and Akeem's presence constituted an interruption in his work flow, which could explain why the request for help was not taken up.

Dynamic representations of context

Video allows for the creation of both specialized static representations (e.g. multimodal transcription, journey maps or activity maps with interactional details inserted) and dynamic representations which can include pathway traces, side-by-side or split-screen video, GIFs, and time lapse video. Time scales can be altered to create representations of video data that can be helpful for analysis. Slowing down time can elucidate what is happening on a micro-level in a short clip and speeding up video clips can help researchers trace trajectories of phenomena over longer periods of time. For example, short video clips collected from multiple classrooms in the STEM study were placed side by side and played in fast motion to display patterns in the children's movement and behavior. Watching the fast motion clips side by side highlighted differences in the activity structure, and allowed us to see how drastically it had changed by the middle of the school year. Representations can be used to get a clearer understanding of the context, and decisions about the most appropriate and effective representations are likely to change as the understanding of the data evolves. In the case of understanding the complex activity systems in our examples, experimenting with different representations has been essential to recognizing patterns and understanding what important things may be missing from the story.

Lessons learned

Capturing visual details makes it possible to consider the context surrounding events and allows for a more complete understanding of the moments of interest. We suggest that researchers who engage in video analysis select different representational forms, play with changes in time scale, and use multiple perspectives to make sense of data, keeping in mind the importance of allowing the whole story to unfold. These manipulations in the representation of context can help focus researchers on salient aspects of context in new ways throughout analysis. Flexibility among representations can allow researchers to consider context in a multi-faceted way.

Conclusions and implications

This symposium contributes to the learning sciences by providing theoretical and practical standards and heuristics for addressing the unique affordances and challenges of the *process* of video analysis, as well as explicating reflexive connections between these theories and methods. These standards and heuristics are at a smaller grain size than overarching themes or guiding principles; they are theoretically-grounded criteria that guide individual analytic decisions. This contribution to learning sciences theory and methodology is timely, as researchers in our discipline increasingly rely on video records to capture and analyze the processes of teaching and learning. It is important, given that the affordances of video and its varied uses in research are consistently outpacing the development of theory and methodology surrounding this medium. Our theoretically-guided set of standards and heuristics, drawn from a diverse group of researchers, research projects, data corpuses, and research questions, contribute to the development of rigorous analytic standards for video research.

References

Barron, B., & Engle, R. A. (2007). Analyzing data derived from video records. *Guidelines for video research in education: Recommendations from an expert panel*, 24-43.

Bezemer, J., & Mavers, D. (2011). Multimodal transcription as academic practice: A social semiotic perspective. *International Journal of Social Research Methodology,* 14(3), 191–206.

Bucholtz, M. (2000). The politics of transcription. *Journal of Pragmatics*, 32(10), 1439-1465.

Derry, S. J., Pea, R. D., Barron, B., Engle, R. A., Erickson, F., Goldman, R., Hall, R. & Sherin, B. L. (2010). Conducting video research in the learning sciences: Guidance on selection, analysis, technology, and ethics. *The Journal of the Learning Sciences*, 19(1), 3-53.

Engle, R. A., Conant, F. R., & Greeno, J. G. (2007). Progressive refinement of hypotheses in video-supported research. In Goldman, R., Pea, R., Barron, B., & Derry, S. J. (2007), *Video Research in the Learning Sciences.* New York: Routledge.

Glaser, B. G. & Strauss, A. L. (1967). The Discovery of Grounded Theory: Strategies for Qualitative Research. Chicago, IL: Aldine Publishing Company.

Goldman, R. (2007). Video representations and the perspectivity framework: Epistemology, ethnography, evaluation, and ethics. In Goldman, R., Pea, R., Barron, B., & Derry, S. J. (2007). *Video research in the learning sciences.* New York: Routledge.

Goldman, R., Erickson, F., Lemke, J., & Derry, S. J. (2007). Selection in video. *Guidelines for video research in education: Recommendations from an expert panel*, 15-22.

Goodwin, C. (2003). Conversational frameworks for the accomplishment of meaning in aphasia. *Conversation and brain damage*, 90-116.

Goodwin, C. (2013). The co-operative, transformative organization of human action and knowledge. *Journal of pragmatics*, *46*(1), 8-23.

Hall, R. (1999). The organization and development of discursive practices for "having a theory". *Discourse Processes*, *27*(2), 187-218.

Hall, R., & Stevens, R. (2016). Interaction Analysis Approaches to Knowledge in Use. In diSessa, A. A., Levin, M., & Brown, N. J. S. (Eds.), *Knowledge and interaction: A synthetic agenda for the learning sciences*. New York, NY: Routledge.

Hutchins, E. (1995). How a cockpit remembers its speeds. Cognitive science, 19(3), 265-288.

Jefferson, G. (2004). *Glossary of transcript symbols with an introduction.* In Lerner, G., *Conversation analysis: Studies from the first generation.* Amsterdam/Philadelphia: John Benjamins Publishing Company.

Jewitt, C. (Ed.). (2009). The Routledge handbook of multimodal analysis.. London: Routledge.

Keifert, D. T. (2015). *Young children participating in inquiry: Moments of joint inquiry and questioning practices at home and in school* (Doctoral dissertation). Retrieved from Proquest. (3724286).

Latour, B. (2005). Reassembling the social-an introduction to actor-network-theory. Oxford University Press. ISBN-10: 0199256047. ISBN-13: 9780199256044, 1.

Lave, J. (1988). Cognition in practice: Mind, mathematics and culture in everyday life. Chicago, IL: Cambridge University Press.

Leinhardt, G., & Crowley, K. (1998). Museum learning as conversational elaboration: a proposal to capture, code, and analyze talk in museums. *Museum Learning Collaborative, http://museumlearning.org/paperresearch.html (Accessed September 1, 2014)*.

Machamer, P., Darden, L., & Craver, C. F. (2000). Thinking about mechanisms. *Philosophy of Science*, 1-25.

McDermott, R. P., Gospodinoff, K., & Aron, J. (1978). Criteria for an ethnographically adequate description of concerted activities and their contexts. *Semiotica*, 24(3-4), 245-276.

Miles, M. B., & Huberman, A. M. (1*994). Qualitative data analysis: An expanded sourcebook.* Sage.

Miles, M. B., Huberman A. M. & Saldana, J. (2014). *Qualitative data analysis: A methods sourcebook* (3rd ed.). Sage.

Ochs, E. (1979). Transcription as theory. In E. Ochs & B. Shieffelin (Eds.), *Developmental Psychology* (pp. 43-72). New York, NY: Academic.

Saxe, G.B. (1999). Cognition, development, and cultural practices. In E. Turiel (Ed.), *Culture and Development. New Directions in Child Psychology.* Jossey-Bass.

Scribner, S. (1984). Studying working intelligence. Cambridge, MA: Harvard University Press:

Stevens, R. (2012). The missing bodies of mathematical thinking and learning have been found. *Journal of the Learning Sciences*, 21(2), 337-346.

Vygotsky, L. S. (1978). Mind in society: The development of higher mental process. Cambridge, MA: Harvard University Press.

Acknowledgements

This material is based upon work supported by the Institute of Education Sciences (U.S. Department of Education R205B080027); the National Science Foundation (grants DRL-1348800, DRL-1433724, SBE-0541957, SMA-0835854, ESI-1020316, and IIS-1123574); the National Science Foundation Graduate Research Fellowship Program (grant DGE-0824162); the AERA-MET Dissertation Fellowship Program; the NAEd/Spencer Dissertation Fellowship Program; and the Institute for Sustainability and Energy at Northwestern University. Contents are solely the responsibility of the authors and do not necessarily represent the official views of the organizations above.