Learning how to look and listen: Building capacity for video-based transcription and analysis in social and educational research

Co-Investigators: Alfredo Artiles Ph.D., Frederick Erickson, Ph.D., Sherman Dorn, Ph.D.

Proposal to the Spencer Foundation.

Proposal Summary. Recently there has been an explosion in technology for video-based research. It is now relatively easy and inexpensive to record with good picture and sound. Yet video-based research faces two enduring challenges: either video is not used at all because new scholars are afraid to try it, or they do not know how to use effectively the visual information the audio-visual record makes available. Until recently there has been very little published on concrete methods of video-based research, and these methods are not nearly so widely taught in universities as are introductory courses in statistics or in general qualitative research methods. Thus beginning scholars who would like to use video as a primary data source may have difficulty in gaining the knowledge and skill they need in order to use video-based analytic approaches in maximally effective ways. This conference provides a way to fill in the current "knowledge gap" in video-based research efforts. A pioneer generation of video researchers and a second generation of researchers will be brought together to demonstrate *how they look at video footage analytically and how they transcribe visual data together with auditory data.* This will be reported in a website and/or video series that can be accessed on the internet. The website will be built and maintained by Arizona State University

CONFERENCE PROPOSAL NARRATIVE

Purpose and Significance. There is a strong theoretical warrant for the notion that social interaction constitutes learning environments (Lave & Wenger, 1991, Rogoff et al. 2001). Fine- grained description and analysis of the real time conduct of interaction, captured in continuous real time audio and video recording, becomes important evidence in the study of teaching and learning in formal and informal learning environments.

New developments in audiovisual recording and analysis technology make it easy and inexpensive to record good picture and sound in studies of classrooms and other teaching/learning settings. But the video record, once made, must be systematically mined as a data source. Although new coding software is being developed, the preparation of detailed transcripts and other data sources concerning verbal and nonverbal aspects of interaction in learning environments requires a human observer who must know how to look and listen to the video recording analytically. Often in the past the video recorder has been treated as an audio recorder—what got research attention was the speech behavior that was recorded rather than the visual information available from the recording. How to look as well as listen analytically— habits of mind in noticing and procedures of analysis based on that noticing—is craft knowledge that can be learned. To demonstrate that craft knowledge concretely and record it for posterity so that new

scholars can learn those skills and develop them further is the purpose of the proposed conference.

Staff at agencies that fund research in the learning sciences (e.g. NSF, the National Academy of Education/Spencer Foundation fellowships program, and the National Institute for Child Health and Human Development) have realized that many research proposals come from early career scholars who have little experience in video-based methods of research and who consequently make beginner's mistakes in analyzing the video footage they collect. Because of this the NSF, the Interagency Education Research Initiative, and the Data Research and Development convened in 2007 a conference of researchers with experience in the use of video as a primary data source, with Derry et al. (2010) appearing afterwards as a special issue of the *Journal of the Learning Sciences*, "Conducting video research in the learning sciences: Guidance on selection, analysis, technology, and ethics." (An earlier special issue of the *Journal of the Learning Sciences*, an issue titled "Interaction Analysis," had focused primarily on issues of video-based data analysis (Jordan & Henderson, 1995).

Over the past 15 years three book length discussions of uses of video in social research have appeared in Europe and the United States: Banks (2001): Knoblauch et al. (2006); Heath, Hindmarsh, and Luff (2010)—and most recently a special issue of the British journal, *International Journal of Social Research Methods* (Jewitt, 2011). [See the overview discussion of the field in Jewitt, 2011, and the brief history of uses of video in social research authored by Frederick Erickson, this proposed conference's Co-Investigator; Erickson, 2011a)].

Early discussions of video analysis of social interaction in the study of teaching and learning were prepared by McDermott, Gospodinoff, and Aron (1978), Erickson and Shultz (1979/republished in 1996), and Mehan (1979). Erickson continued to publish chapter length discussions of video-based research methods in education: Erickson 1982, 1992, 2006, and 2007, including broader interdisciplinary reviews of the history of the use of audiovisual recordings in social research (Erickson, 2004, 2011a, 2011b). Erickson was also a co-author of the article on video data selection and analysis that was published in the special issue discussed above, edited by Derry (Derry et al., 2010).
Only recently has the literature on methods of video analysis begun to expand. And the systematic use of audiovisual records to study social interaction is only about sixty years old. The first audiotape that recorded casual conversation for research purposes was made in 1953 (see Soskin & John, 1963) and the first systematic analysis of interaction using sound cinema film was done in 1956-57 by a research group in the first full-year class at the Center for Advanced Study in the Behavioral Sciences at Stanford (see McQuown, 1971, in an unpublished report titled "The Natural History of an Interview" [NHI]).

What resulted from the pioneering efforts of that research group was an approach that came to be called "context analysis," a term used in further developments of the NHI approach by Scheflen (1973) and Kendon (1990). The overall NHI perspective also influenced a next generation of younger scholars (notably McDermott et al., 1979; Mehan, 1979; Erickson & Shultz, 1982; C. Goodwin, 1994; M. Goodwin, 2006; Jordan &

Henderson 1995; Streeck, 1983; Streeck & Mehus 2004; Streeck, Goodwin, & LeBaron, 2010), some of whom called this approach "microethnography."

The NHI-derived approach presumed that face-to-face communication was a *semiotic ecology*, consisting in relationships of mutual influence between what speakers and listeners were doing, verbally and nonverbally, through their continuous activity in real time. (The ecology is "semiotic" in that it involves the exchange of the full range of kinds of communicative signs and expressive media—in speech, in non-verbal behavior, in written symbols, and in various types of graphic displays.)
The relevance of this perspective for studies of teaching and learning is that semiotic ecologies, enacted in embodied ways in everyday social interaction, can be considered as *learning environments.* Social interaction, as an ecosystem enacted in real time, is what "communities of practice" do (see Lave & Wenger, 1991; Rogoff et al., 2001), within which novices can develop capacities for more and more central participation over time. As humans we learn across the entire life span in "multi-modal," "embodied," multiply semiotic situations of communication in social interaction—not just learning by listening to words, or by looking at visual displays, but by listening, looking, picking up objects and feeling their weight in motion, speaking, attending to writing systems (e.g. for language and for mathematics) and to systems of graphic representation.

There have been four main approaches to data analysis in video-based studies of teaching and learning (see the discussions in Erickson, 2006, 2011; and in Jewitt, 2011). The first approach involves the use of deductively constructed coding schemes, in effect, checklists of function-categories, usually focusing on verbal instructional "moves" by the teacher or by students. The function-categories coding approach derived from the early work of Bales (1950). It was employed extensively in the "process-product" approach to the study of teacher effectiveness (e.g., Brophy & Good, 1978). The checklist coding approach has reappeared recently, e.g. in the work of the international mathematics teaching comparisons in the TIMMS video project (see Stigler, Gallimore, & Hiebert 2000) and in current attempts to identify measurable features of pedagogy that correlate with measures of student achievement (see for example, Mashburn et al., 2008).

A second major approach is that of "conversation analysis" (CA). Originating in sociology, this approach focuses primarily (but not exclusively) on certain features of speech (e.g. turn-taking patterns, self- and other-correction). For an example, see Macbeth (2004).

A third major approach derives from qualitative research traditions in anthropology and sociology, and sociolinguistic discourse analysis. This approach combines firsthand participant observation, research subject interviewing, and collection of site documents with video recording and analysis. (Much of this work has been influenced by the "context analysis" perspective reviewed above, although in many cases that influence is indirect rather than direct.)

A fourth approach involves analysis of classroom video records by subject matter specialists (e.g., Roth, 2005). Without a background in mainstream discourse analysis and sociolinguistics, subject matter specialists develop coding schemes and other analytic

procedures inductively, and conduct analysis that vary considerably from one scholar to the next.

The preponderance of research on classroom discourse as a learning environment over the past 25 years has taken place along the lines of the third approach reviewed immediately above, and that is also the perspective taken in early and more recent publications on video-based research methods. Accordingly, the proposed conference would emphasize primarily the third approach, sociolinguistic/ethnographic discourse analysis, with some representation from conversation analysis, and some from subject matter specialists.

Overview of What Will Be Learned. The conference will bring together an interdisciplinary group of older and younger scholars in education, anthropology, sociology, linguistics, and psychology who have done exemplary studies of face-to-face interaction using video that take particular advantage of visual information available in video recording. Conference activities will be designed to reveal the craft knowledge of these scholars—how, concretely, using continuously shot and unedited video records, they look at the video track analytically as well as listen to the audio track. Conference recordings will capture that craft knowledge for public distribution and availability.

Conference participants will share their analysis and transcription procedures in four ways, providing material for the preparation of a written report and a website/video series that will highlight similarities and differences in the participants' analytic approaches. Participants will share by (1) preparing a brief reflection paper (no more than ten pages) with video illustration, based on previous research, in which *how they looked analytically at the footage* will be the primary focus of discussion. These reflections will be presented in plenary sessions during the conference; (2) reviewing a two-minute video clip showing the teaching of a key idea in the physics of matter—that matter occupies space--in a kindergarten-first grade classroom. As the conference begins each participant will be video recorded individually, viewing this video clip for a half hour, talking aloud about what they notice while playing the tape, stopping and replaying, as they conduct the first steps in an analysis and transcription of the information they attend to while viewing the video clip; (3) discussing in a plenary "viewing session" in which all the conference participants will watch the video clip together, comment on what they notice, and discuss differing approaches to initial analysis; (4) participating in a concluding plenary discussion followed by one-page, individually-written reflections in which conference participants will comment on ways in which differences and similarities in their ways of approaching initial data analysis—their ways of looking and listening analytically with video footage—can be communicated in a written conference report, as well as in a website or video series using conference footage. All plenary sessions of the conference will be video-recorded, as well as the individual viewing sessions discussed in (2) above, providing resources for preparation of the conference products discussed in the next section.

Expected Outcomes and Products. (1) A written report of the conference, summarizing the individual video clip viewing sessions as case studies of beginning analysis practice, and also summarizing the whole-group viewing session discussion and the concluding

plenary discussion. The report will also include all reflection papers prepared and delivered by conference participants. Publications in journals may follow from this report. (2) In a website, reflection papers prepared by each of the conference participants (and, pending IRB permission, the video examples for each reflection paper) will be available for reference, together with the written conference report. (3) Also in the website, or in a separate video series (which of these dissemination formats is most appropriate and feasible will be determined after the conference takes place), examples of analysis approaches will be presented, using the video footage from individual scholars' video clip viewing sessions, and from the plenary discussions. These will be audiovisual case studies of differing approaches to data identification, analysis, and transcription showing how different scholars approach this concretely, in their looking while listening.

Conference products will document the current "state of the art" in ethnographic approaches to the study of social interaction as a learning environment—the craft analytic knowledge of a pioneer generation and of a successor generation of scholars. The conference products will also highlight differences and similarities in the participants' theoretical orientations, units of analysis, and approaches to data representation. Most fundamentally, the written report and website, built and maintained by ASU will be a resource for future instruction in the primary operation of *looking analytically while listening*, skills and perspectives in noticing that are foundational for video-based studies of social interaction as a learning environment.

REFERENCES

Bales, R. (1950). *Interaction process analysis: A method for the study of small groups.* Cambridge, MA: Addison-Wesley.

Banks, M. (2001). *Visual methods in social research*. London: Sage.

Brophy, J., & Good, T. (1978). *Looking in classrooms.* New York, NY: Harper and Row.

Derry, S., Pea, R., Barron, B., Engle, R., Erickson, F., Goldman, R., Hall, R., Koschmann, T., Lemke, J., Sherin, M., & Sherin, B. (2010). Conducting video research in the learning sciences: Guidance on selection, analysis, technology, and ethics. *Journal of the Learning Sciences, 19*(1), 3-53. (A white paper by an expert panel, prepared for the National Science Foundation, the Interagency Education Research Initiative, and the Date Research and Development Center)

Erickson, F. (1982). Audiovisual records as a primary data source. *Sociological Methods and Research, 11*(2), 213–232.

Erickson F. (1992). Ethnographic microanalysis of interaction. In M. LeCompte, W. Millroy, & J. Preissle (Eds.), *The Handbook of Qualitative Research in Education* (pp. 201-225). Orlando, FL: Academic Press, Inc.

Erickson, F. (2004). Origins: A brief intellectual and technological history of the emergence of multimodal discourse analysis. In P. Levine & R. Scollon (Eds.), *Discourse*

and technology: Multimodal discourse analysis (pp. 196-207). Washington, DC: Georgetown University Press.
Erickson, F. (2006). Definition and analysis of data from videotape: Some research procedures and their rationales. In J. Green, G. Camilli, P. Elmore, A. Skukauskaite, & E. Grace (Eds.), *Handbook of complementary methods in educational research* (pp. 177-192). Washington, DC: American Educational Research Association.

Erickson, F. (2007). Ways of seeing video: Toward a phenomenology of minimally edited footage. In R. Goldman, Barron, R. Pea, & S. Derry (Eds.), *Video research in the learning sciences (*pp. 145-155). Mahwah, NJ: Lawrence Erlbaum Associates.

Erickson, F. (2011a). Reflections on practice, teaching/learning, video, and theorizing. In T. Koschmann (Ed.), *Theories of learning and studies of instructional practice* (pp. 385- 402). New York, NY: Springer.

Erickson, F. (2011b). Uses of video in social research: A brief history. *International Journal of Social Research Methodology, 14*(3), 179-189.

Erickson, F., & Shultz, J. (1979/1997). When is a context? Some issues and methods in the analysis of social competence. In M. Cole, Y. Engeström, & O. Vasquez (Eds.), *Mind, culture, and activity: Seminal papers from the Laboratory of Comparative Human Cognition* (pp. 22-31). New York, NY: Cambridge University Press. (Originally published in 1979 in the Newsletter of the Laboratory of Comparative Human Cognition).

Goodwin, C. (1994). Professional vision. *American Anthropologist*, *96*(3), 606–633.

Goodwin, M. (2006). *The hidden life of girls: Games of stance, status, and exclusion.* Oxford: Blackwell.

Hall, R., Ma, J., & Nemirovsky, R. (2015). Re-scaling bodies in/as representational instruments in GPS drawing. In V. Lee (Ed.), *Learning technologies and the body: Integration and implementation in formal and informal learning environments* (pp. 112-131). New York, NY: Routledge.

Heath, C., Hindmarsh, J., & Luff, P. (2010). *Video in qualitative research: Analysing social interaction in everyday life*. London: Sage.

Jewitt, C. (2011). Editorial [Introduction to a special issue on uses of video in social research]. *International Journal of Social Research Methods, 14*(3), 171-178.

Jordan, B., & Henderson, A. (1995). Interaction analysis: Foundations and practice. *The Journal of the Learning Sciences, 4*, 39-103.

Haviland, J. (2010). Musical spaces. In C. Streeck, C., Goodwin, & C. LeBaron (Eds.), *Embodied interaction: Language and body in the material world*, (pp. 299-304). Cambridge: Cambridge University Press.

Kendon, A. (1990). *Conducting interaction: Patterns of behavior in focused encounters.* New York, NY: Cambridge University Press.

Knoblauch, H., Schnettler, B., Raab, J., & Soeffner, H. (Eds.). (2006). *Video analysis – methodology and methods: Qualitative audiovisual data analysis in sociology*. Frankfurt: Peter Lang.

Kress, G., Charalampis, T., Jewitt, C., & Ogborn, J. (2001). *Multimodal teaching and learning: The rhetorics of the the science classroom.* London: Continuum.

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation.* Cambridge: Cambridge University Press.

Macbeth, D. (2004). The relevance of repair for classroom correction. *Language in Society, 33*, 703-736.

Mashburn, A., Pianta, R., Hamre, B., Downer, J., Barbarin, O. Bryant, D., Burchinal, M. Clifford, R., Early, D., & Howes, C. (2008). Measures of classroom quality in pre-kindergarten and children's development of academic, language, and social skills. *Child Development, 79*, 732-749.

McDermott, R., Gospodinoff, K., & Aron, J. (1978). Criteria for an ethnographically adequate description of concerted activities and their contexts. *Semiotica, 24*, 245-275.

McQuown, N. (Ed.) (1971). *The natural history of an interview.* University of Chicago, Joseph Regenstein Library, Microfilm department, Microfilm Collection of Manuscripts on Cultural Anthropology, Series 15, Nos. 95-98.

Mehan, H. (1979). *Learning Lessons: Social organization in the classroom.* Cambridge, MA: Harvard University Press.

Rogoff, B., Turkanish, C., & Bartlett, L. (2001). *Learning together: Children and adults in a school community.* New York, NY: Oxford University Press.

Roth, W.-M. (2005). *Talking science: Language and learning in science.* Lanham, MD: Rowman and Littlefield.

Scheflen, A. (1973). *Communicational structure: Analysis of a psychotherapy transaction.* Bloomington, IN: University of Indiana Press.

Soskin, W., & John, V. (1963). The study of spontaneous talk. In R. Barker, & L. Barker (Eds.), *The stream of behavior* (pp. 228-281). New York, NY: Appleton-Century-Crofts. (Audio recording made in 1953.)

Stigler , J. W., Gallimore, R., & Hiebert, J. (2000). Using video surveys to compare classrooms and teaching across cultures: Examples and lessons from the TIMSS and TIMSS-R video studies. *Educational Psychologist, 35*(2), 87-100.

Streeck, J. (1983). *Social order in child communication: Pragmatics and beyond.* Amsterdam: Benjamins.

Streeck, J., & Mehus, S. (2004). Microethnography: The study of practices. In K. Fitch, R. Sanders, J. Bavelas, & R. Arondale (Eds.), *Handbook of language and social interaction* (pp. 381-404). Mahwah, NJ: Lawrence Erlbaum Associates.

Streeck, J., Goodwin, C., & LeBaron, C. (2010). *Embodied interaction: Language and body in the material world.* Cambridge: Cambridge University Press.