The Process of Creation:
A Novel Methodology for Analyzing Multimodal Data

Erica Rosenfeld Halverson, Michelle Bass, and David Woods
University of Wisconsin-Madison, Madison, WI, USA

In the 21st century, meaning making is a multimodal act; we communicate what we know and how we know it using much more than printed text on a blank page. As a result, qualitative researchers need new methodologies, methods, and tools for working with the complex artifacts that our research subjects produce. In this article we describe the co-development of an analytic methodology and a tool for working with youth produced films as multimodal artifacts of youth engagement with identity. Specifically, we describe how to employ this multimodal framework in data analysis, with an emphasis on how different modes interact with one another, and how new meanings are made possible through multimodal interactions. Key Words: Multimodality, Data Analysis, Video Production, Youth Development, Qualitative Research.

Hard cut. Moving on a road past palm trees in Jamaica. Close up of posters, one that says One Love and has a picture of Bob Marley. Film of Jamaican man and little girl, who at first looks at the camera but then turns away.
Voiceover: So along with a new mom, came siblings, cousins, aunts, uncles, and friends.
Family climbing up a waterfall, one little white girl, Melinda is at the top of the line.
Voiceover: "We had fun in Jamaica, we had fun in New York".
Panning Tree line, fruit, chicken coop, chickens under the coop, someone swimming in the water.
Bob Marley music: I wanna love ya.
Voiceover: But the fun didn't last.
Bob Marley music: And treat you right.
Voiceover: When I was eight years old.
Bob Marley music: I wanna love ya.
Voiceover: Beverly and Nikki moved to Florida.

The above text is the transcript of a clip from *Jewmaican*, a short documentary film produced at Reel Works Teen Filmmaking. It is a film about a young woman whose mother committed suicide when she was an infant. She was raised in part by her Jamaican nanny who, "ultimately became her mom" (Reel Works, n.d.). The film explores how she reconciles her experiences with her biological, Jewish family and her adopted, Jamaican family. We came across Jewmaican through our research on how young people learn to make digital art about the stories of their lives through their participation in youth media arts organizations (Halverson & Gibbons, 2010; Halverson, Lowenhaupt, Gibbons, & Bass, 2009). In our work, we have been interested in

understanding both how the artistic production process happens and what the products represent. Specifically, we are interested in how filmmaking supports young people in exploring and representing complex issues of identity through the construction of multimodal artifacts. As the above transcript hopefully makes apparent, we found traditional methods of engaging with discourse – the creation of a flat text-based transcript and discourse analysis – unsatisfactory for capturing films as products of identity. Our dissatisfaction led us to a fundamental methodological question: How can we analyze youth produced films as multimodal artifacts of youth engagement with identity?

This article represents our nearly two-year journey developing a successful and reproducible method for analyzing multimodal data. Our work is grounded in multimodality as a theoretical lens to inform the co-construction of an analytic methodology and an analytic tool for working with multimodal, video-based data. We draw specifically on Günter Kress' (2000, 2003; Kress & van Leeuwen, 2006) construction of multimodality coupled with Jay Lemke's (1998, 2002, 2007) understanding of the combinatorial nature of modes in meaning making to frame our analysis of youth produced videos as multimodal texts. Using research that explores how youth-produced videos provide evidence of identity exploration and representation in action, we outline the development of a multimodal methodology inspired by our desire to understand the products of youth filmmaking. We then describe the co-evolution of this analytic methodology and video analysis software tool that focuses on the multimodal nature of video data, particularly ways of looking at the interaction within and across modes. Finally, we offer other examples of how our model for multimodal data collection and analysis can extend current educational research practices that use video data to understand meaning making in action.

While our focus is on analyzing youth films as representations of identity, it is important to note that the films themselves are the final products of many months of workshops, shooting, editing, and mentoring. In other writing, we describe the results of our instrumental case studies (Stake, 2000) with four youth media arts organizations (YMAOs) across the United States (Halverson, in press; Halverson et al., 2009; Halverson & Gibbons, 2010). Through interviews with organizational leaders and youth, observations of the filmmaking process and the collection of artifacts that trace the development of films over time, we have built a nuanced picture of how youth learn to make films about the stories of their lives and the ways in which this process is similar and unique across organizations. We have identified "key moments" in the film production process across all the organizations that serve as formative checkpoints where youth demonstrate their understanding of the relationship between what they want to signify (in this case personal identity) and how the tools of film can be used to represent identity (Halverson & Gibbons, 2010). While it could be argued that analyzing youth films absent an accompanying analysis of the filmmaking process is incomplete, an understanding of how youth learn to make films is incomplete without a method for analyzing how their learning process is realized in the films they produce. Perhaps more importantly for this article, as standalone data the films demonstrate at a broader scale the ways adolescents from marginalized groups use the multimodal resources of film to actively construct and represent identity.

Theoretical Framework, Methodology, and Method: A Brief Discussion

Since we aim to describe and demonstrate a novel method, methodology, and tool for working with multimodal data, it is important to be clear about what we mean by these terms. We draw on Crotty's (2003) organization for the foundations of qualitative social science research, summarized in Table One.

Table 1. *Summary of Four Primary Elements of Qualitative Research*

Area	Question	Description
Method	What methods do we propose to use?	Concrete techniques or procedures.
Methodology to shape	What methodology governs our choice and use of methods?	Strategies and plans choice of methods.
Theoretical Framework	What theoretical perspective (framework) lies behind the methodology in question?	Philosophical stance that underlies choice of methodology.
Epistemology	What epistemology informs this theoretical perspective?	Theories of Knowledge.

Oftentimes, these four terms are used interchangeably, incorrectly, or one or more questions are ignored in the research process. Terminology in qualitative research can make complicated work even more tangled. Whereas methods refer to specific techniques and procedures such as interviews, field notes, and surveys, methodologies capture the strategies and plans of action that shape "our choice and use of particular methods and links them to desired outcomes" (Crotty, 2003, p. 7), such as conducting an ethnography, or deciding between a narrative or discourse analysis. A theoretical framework is the philosophical stance "that lies behind our chosen methodology" (Crotty, 2003, p. 7). It is used to provide context for the process. Therefore, the theoretical framework informs the methodological choices that inform method selection, which ultimately informs the types of data collected and the types of analyses that can be performed on the data. While we will not spend much time describing our epistemological stance, our work sits comfortably in a constructionist perspective, where the subjects and objects of research are partners and truth is constructed through their interactions and relationships (Crotty, 2003).

The Importance of Multimodality as a Framework and a Methodology for Analysis

Multimodality has been described as both a theoretical framework and an analytic methodology for understanding how people make meaning using the sign systems that are available to them. As a theoretical frame, multimodality allows us to construct the task of meaning making in terms of the semiotic resources available (Kress, 2000, 2003; Kress & van Leeuwen, 2006; Lemke, 1998). As Kress (2000) describes, "the assumption underlying a multimodal approach to communication and representation is that…humans use many means made available in their cultures for representation precisely because these offer differing potentials, both for representation and for communication" (p. 194). Mitchell (1994) describes the *pictorial turn* in society as a shift from identifying meaning making resources as primarily textual to the importance of image/text relationships. Specifically, the emergence of new digital technologies has created new forms of meaning production that are accessible worldwide. The pictorial turn represents, "a postlinguistic, postsemiotic rediscovery of the picture as complex interplay between visuality, apparatus, institutions, discourse, bodies, and figurality" (p. 16). A multimodal framework attends to the pictorial turn through an understanding that people make meaning using these myriad resources.

As an analytic methodology, multimodality allows us to understand how people use their meaning making resources in context (Eggins, 1994; Kress & van Leeuwen, 2006; LeVine & Scollon, 2004). A multimodal analysis incorporates all the communicative modes that can be identified in the scope of recorded human interaction (Norris, 2004) allowing researchers to answer both the question of how people use their linguistic resources and how these resources are structured for use (Eggins, 1994). In working with multimodal discourse analysis as methodology, "it becomes apparent that [the various modes] are intricately interwoven, they are not easily separable, and they are interlinked and often interdependent" (Norris, 2004, p. 102). Multimodal analysis, then, is much more than additional layers of analysis but rather a consideration for how the layers work together to create new meanings.

The development of analytic methodologies, methods and tools that attend to multimodality is timely given the emergence of video research as an epistemology, a method, and a tool for collecting and analyzing data. The learning sciences as a field is particularly concerned with how to collect, analyze, and share video data around teaching and learning (Goldman, Pea, Barron, & Derry, 2007). Specifically, Lemke (2007) discusses how we make meaning over time with video data, particularly what and how we attend to a multimedia data source, and Goldman (2007) focuses on the development of analytic criteria, "that take into consideration the range of both evaluative measures and e-value-ative qualities for adjudicating the significance of research using video as a research tool" (p. 6). This work focuses on important methodological questions such as how data are framed and what is gained and lost through the use of video. However, as Mondada (2006) notes, "analytical studies focusing on video as a timed accomplishment and as a social practice are still very few" (p. 51). The alignment around multimodality that we propose may help address some of these issues.

Digital Media as Multimodality

In our own work, we are interested in how multimodality both as framework and as methodology helps us to better understand digital media spaces such as youth media arts organizations (YMAOs). These spaces are compelling participatory environments for youth because they afford the opportunity to engage in the trying on and representing of multiple identities over time (Gee, 2003; Jenkins, Purushotma, Clinton, Weigel, & Robison, 2007; Willett, Burn, & Buckingham, 2005) and it is important to understand how youth become members of these communities and how their membership allows them to explore issues of identity over time. This is especially crucial for youth who feel marginalized in mainstream institutions and who do not have opportunities to explore a positive sense of self in traditional institutional contexts. Understanding how the construction of multimodal representations supports identity development processes can help us to bring these new media literacy practices to youth who are most in need of alternative mechanisms for engaging in positive identity work.

Digital videos (or films) are a specific type of multimodal text. Lemke (2007) describes video and film as, "shar[ing] substantially the same audio-visual semiotic; the same interpretative conventions for their salient sensory features" (p. 41). Manovich (2002), on the other hand, distinguishes between these two media, arguing that the computer-mediated nature of video represents a new semiotic absent in the production of films. Following Lemke's interpretation of the shared semiotics, we use these terms interchangeably. Jay Lemke (2010) describes multimodal media such as digital videos as representations that produce meaning by, "intersecting the semiotic resources of language, visual display, sound and music, cinematic movement, material artifacts, and abstract animation" (n.p.). In our work, we strive to understand how youth construct and represent their personal identities through the multimodal, filmic texts they produce.

Youth are engaged in the production of a wide variety of multimodal, new media artifacts including digital stories, video games, video "mash ups", and spoken word digital poetry, each of which requires the use of a different set of semiotic resources for meaning making. Just as interpreting everyday interactions requires an understanding of how communicative resources work in combination (Norris, 2004), Lemke (2002) argues that the production of multimodal artifacts requires more than understanding how the producer works with each individual mode. Rather, "the meaning potential, the meaning-resource capacity of multi-modal constructs is the logical product, in a multiplicative sense, of the capacities of the constituent resources systems" (p. 303). Since each is unique, the analysis of a digital media product requires attention to the specific semiotic resources involved in the construction of one type of multimodal, digital artifact. The time seems right to consider the analysis of films as mechanisms for young people to make meaning as our society moves from the use of the printed word as the language of cultural communication to the language of audiovisual moving images (Manovich, 2002).

Research on digital story production has demonstrated how youth develop agency by merging words, rhythm, rhyme, music message and image to create personal narratives (Hull & Katz, 2006; Hull & Nelson, 2005; Nelson, Hull, & Roche-Smith, 2008). This work on digital story resulted in a key theoretical insight about the nature of multimodal composition:

> Although different semiotic modes may seem to encode the same content, they are nonetheless conveyors of qualitatively different kinds of messages. The point is that images, written text, music, and so on each respectively impart certain kinds of meanings more easily and naturally than others. (Hull & Nelson, 2005, p. 229)

Unlike digital stories, film as a medium of expression adds an additional communicative tool to this mix – movement. The relationship between the elements described in the film and video work becomes dynamic in both place and time, giving producers an additional tool for representing self (Baldry, 2006; Baldry & Thibault, 2006; Kress & van Leeuwen, 2006; Manovich, 2002). Baldry (2006) argues that researchers need to attend to movement in film as a core tool for meaning making:

> Moreover, we need to access texts in an in vivo form that provides access to audio and video tracks and maintains their relationship intact, because a major part of the way in which a film text makes its meaning is precisely through the synchronization between visual and audio resources. (p. 180)

The kineikonic mode. A key feature of multimodal production and film in particular is that producers strive to communicate meaning, and audiences make meaning, not just with individual modes but also in the ways that modes interact with one another and what is created as a result of their interaction (Burn & Durran, 2006; Burn & Parker, 2003; Kress, 2003; Kress & van Leeuwen, 2006; Mitchell, 1994; Nelson et al., 2008). This is captured in Lemke's (2002) questions surrounding the field of multimodal semiotics: "How do the meanings of multimodal complexes differ from the default meanings of their monomodal components in isolation? How do we construe the meanings of components in multimodal complexes and of whole complexes as such?" (p. 303).

Burn and Parker (2003) make this interaction concrete, calling it the kineikonic mode, "literally, the mode of the moving image" (p. 13). In their media education work with youth, Burn and Durran (2006) use the kineikonic mode to describe how youth come to understand the function of the different modalities of film in representing ideas as they re-edit already existing films. They describe the kineikonic mode as, "combin[ing] a range of different signifying systems, the important ones here being music, visual dramatic sequences, and the affordances of editing – shot structure, transitions, duration, pace, and rhythm" (p. 279). The kineikonic mode describes a new mode that is created in the interaction among two or more modes and is an important concept in the analysis of multimedia (Lemke, 1998) or hyper modal (Lemke, 2002) texts, especially in analyzing youth-produced work (Burn & Durran, 2006; Burn & Parker, 2003; Curwood & Gibbons, 2009).

Moving away from the prioritization of oral production. Kress (2003) describes modern literacy as a shift from the telling the world to showing the world. Much like Mitchell's (1994) pictorial turn, Kress captures an emerging emphasis on visual images as tools for meaning making that are much more than a supplement to verbal tools. Though peoples' words (whether oral or textual) give us significant insights

into meaning, a shift to multimodal analysis encourages us to relinquish our reliance on the spoken word as the dominant mode for meaning. In the analysis of films, it is easy to favor the sound mode, with a heavy focus on dialogue to serve as a natural marker for analytic segments, marginalizing the other three modes as afterthoughts. However, when faced with a film in our data set that did not use any dialogue, we began to explore the weight of non-verbal modalities in meaning making. In fact, our work with this film marked the beginning of our journey toward a new way of thinking about a multimodal framework, methodology, and method.

*Skin***.** The beginning of the process for designing a multimodal analytic methodology began in the fall of 2007 with a non-conventional youth produced film, *Skin*. *Skin* was produced in 2005 by several youth from the TRUCE Arts and Media program at the Harlem Children's Zone:

> *Skin* is an experimental video piece that explores the issues of discrimination, racial identity and self-esteem. It follows two young African Americans who attempt to change the color of their skin and are haunted by the effects of their decision. Ultimately, they come to rediscover themselves, and the rich natural beauty of their own skin. (ListenUp!, n.d. b)

Our initial approach to film analysis was to begin with oral modalities, and dialogue in particular. Encountering a film with no dialogue meant we had to rethink our approach. We turned to Baldry and Thibault's (2006) social semiotic approach to analysis for guidance in how to start our analysis of the meaning making process in this "silent" film. We broke the film into phases and transitions, units of analysis that demonstrate "semiotic homogeneity", internal consistency across multiple modes, for example, the same music, voiceover, and shot type (Baldry & Thibault, 2006). We then developed a coding scheme based on Bordwell and Thompson's (2004) formal analysis of films to populate the phase and transition units. In formal analysis, films are interpreted based on the four key cinematic techniques employed by filmmakers:

- · Mis-en-scene. Anything within the frame of the camera including subject-related elements, setting, scripted features, and style.
- · Sound. Anything you hear in the film, specifically, dialogue, sound effects, and music.
- · Editing. The work the filmmaker does after shooting is completed in order to assemble the film.
- · Cinematography. Techniques used to alter the image seen through the camera lens.

We used these four cinematic techniques as the four primary modes of representation that are analyzed with the phases and transitions of a film. This is consistent with Burn and Durran's (2006) assertion that music, visual dramatic sequences and editing serve as the primary mechanisms for the creation of the kineikonic mode. Our addition of

cinematography highlights our emphasis on original production, unlike Burn and Durran's use of already existing footage in their work with youth.

The four primary modes represent the broad categories by which we classified the "filmic elements" useful for analysis. Within each of these modes are sub-codes that include instantiations of each element. For example, the sound element includes codes for content including dialogue, sound effects, and music as well as codes for the origin of the sound including diegetic, non-diegetic, and internal diegetic (see Appendix A for full list of codes). The YMAOs in our study provided instruction around the use of these elements and the variety of meanings they convey depending on their use. Using these categories to describe the phases and transitions of the films resulted in the creation of multilayered filmic transcripts that allow us to consider each mode individually, as well as how they connect to one another to help youth consider issues of identity in their films.

Our initial use of this method was decidedly non-technological. Using a series of index cards, the six members of our research team wrote down what we saw on the screen. We began with the mis-en-scene elements of the film. When we reached consensus across at least four team members, we entered these visuals into an excel file and created linear time codes that allowed us to distinguish phases, transitions, and their sub-units. What we saw on the screen when we created this initial representation was similar to current video research that explores gestural analyses of meaning making (Alibali & Nathan, 2007; Barron, 2007; Singer, Radinsky, & Goldman, 2008). Just as textual narrative can provide the initial road map for interpretation, the mis-en-scene transcript allowed us to see the overall flow of the film. With the mis-en-scene as our guide, we turned to the sound transcript, which features an instrumental version of *Feelin' Good*, a song originally written for the Broadway musical *The Roar of the Greasepaint – the Smell of the Crowd*, but popularized by Nina Simone (Wikipedia, n.d.). While there are no lyrics in the film, those who are familiar with the song will recognize that at the moment when the two actors choose to remove the paint they have spread on their faces to cover their true skin color, the lyrics would read: "It's a new dawn, it's a new day, it's a new life. And I'm feelin' good". Not only do those implied lyrics demonstrate the hopefulness present in the actors' actions and expressions, the music itself shifts in tone from mournful to uplifting.

Our initial method captured the relationship between mis-en-scene and music in a way that allowed us to "see" the kineikonic mode in action and to write about it. However, films are more than just sound and movement. The *Skin* filmmakers also used editing to great effect. Throughout the film, they use three different editing techniques: quick, hard cuts, split screen, and a slow fade from black and white to color. The film is constructed as a contrast between two subjects: a light-skinned, African-American young man and a dark-skinned, African-American young woman. Both are dissatisfied with their skin color, as evidenced by their looking in the mirror and making the decision to paint their faces with their opposite color. The quick, hard cuts that dominate most of the film remind the audience of the opposite, yet parallel lives these young people are leading, culminating in a split screen that appears as if they are looking at (or past) one another. In the penultimate phase of the film, marked sound-wise by the shift to uplifting music described above, both young people scrub the paint from their faces and the screen fades slowly from black and white to color. We found ourselves complicating the visual

diagram of mis-en-scene and sound by inserting descriptions of editing choices in order to include this component of meaning into our interpretation.

Finally, while there is no dialogue in this film, there is limited text in the final phase of the film. The phrases "I AM PROUD," and "BLACK IS WHO I AM" appear one after the other in a quick fade, as the coda to the piece. The use of the pronoun "I" is particularly interesting. There are two actors in the film so the "I" can be interpreted as referring to each of them individually or to a more collective "I", the community of African American young people who they represent. The limited use of text in this film conveys a strong message about identity as both an individual and a collective phenomenon in the filmmakers' lives. While we included this in part as a mis-en-scene element, its unique status as text was not conveyed in our linear, excel-based representation. While we felt good about our evolving method, we needed a tool that adequately captured how we were situating our method in a multimodal framework and methodology.

Transana as a Tool for Multimodal Analysis

Skin was a defining film in understanding the importance of multimodality and the necessity of creating multiple, simultaneous transcripts for conducting analysis; we needed a tool that would make this type of analysis possible. We turned to Transana, a software tool for the transcription and qualitative analysis of video and audio data (www.transana.org). Transana's primary features include the capacity to:

· Transcribe data
· Identify analytically interesting clips
· Manipulate clips by assigning key words, arranging and rearranging, creating complex collections of interrelated clips
· Explore relationships between applied keywords
· View graphical and text-based reports about analytic coding
· Share analysis with colleagues

Transana is used by researchers around the world in a wide variety of disciplines, including computer science (Kola, Kosar, & Livny, 2004), health (Probst, DeAgnoli, Batterham, & Tapsell, 2009), education research (e.g., Halverson, 2010; Mavrou, Douglas, & Lewis, 2007; Alibali & Nathan, 2007), as well as in sociology, psychology, economics, business, medicine, and law. Subjects as varied as human political discourse, learning during video game play, and reptilian social behavior have been studied using the software. It is difficult to track the use of Transana since it is an open source project, though it is currently in use in over 30 countries and is available in 10 languages.

There are a number of computer-assisted qualitative data analysis (CAQDAS) tools that support the analysis of video, most notably NVivo8, DIVER, and ORION (for these and others, see Pea & Hoffert, 2007, pp. 446-447). All of these tools could support a multimodal analysis of the kind we describe, though each tool has its own focus for analysis based on the goals of the developers. NVivo8, for example, is an extension of an analytic tool designed to sort, manage, and restructure large bodies of qualitative data. Its primary affordances include the creation of video clips and the ability to link between the

clips and the source video as an analogue to the original NVivo, designed to support ways to manipulate textual data (NVivo8, 2007). NVivo is a continually evolving tool, with new versions being produced as responses to user's requests for new functions and the increased capabilities that arise from technological advances.

Other popular qualitative software for analyzing video data worth mentioning include DIVER and ORION. DIVER affords multiple analysts the opportunity to take different points of view on the same video-based interaction. Each analyst can develop a different, complimentary analytic orientation toward the same interaction, developing his or her own, "dive" or "take" on the data (Zahn et al., 2005). ORION was created as a tool for video-based, ethnographic research across multiple users and, "designed for use by a community of researchers, teachers, and learners in distributed locations to make meaning of rich video data" (Goldman, Crosby, Swan, & Shea, 2005, p. 114). ORION affords extensive commenting on and clustering of video clips, which allows analysts to consider the relationship among and across video clips, building meaning over time.

All of these tools are based on an analytic frame that focuses on the content of the videos – whether the analysis is socially constructed (as in ORION) or based on a reframing of the data itself (as in DIVER), the data of interest are in the content of the video. In our case, the data of interest are both the content of the video (what is in the films) and how the content is presented (the choices filmmakers made in representing what is in the films). None of the tools available to us at the time we started our project afforded a focus on the different modalities used in production and how these modalities worked independently and together to create meaning. Our co-location with the developers of Transana allowed us to work with them to modify the design of the tool to meet our analytic needs.

Features and Affordances of the Multimodal Analytic Methodology and Tool, Transana 2.3

In this section, we highlight some of the key features of our methodological approach and the accompanying features of Transana that operationalize our approach. We want to make two, interrelated points: (a) Analysis of the kineikonic mode requires the design of an analytic tool that allows for multiple, simultaneous transcripts and; (b) In building a framework and tool that attends to the kineikonic mode, we came to understand the relationship between youth media arts production and identity.

Multiple, Simultaneous Transcripts

When we began analyzing the corpus of films in our data set, we worked with Transana 2.2. In this version, there was a single window provided for transcripts of video data that was then linked to the video itself. In constructing the textual transcripts that link to the video, however, it became clear that a single transcript for textual representation was insufficient. As we described earlier, our coding method requires we attend to four modes of analysis – mis-en-scene, sound, editing, and cinematography – as well as the kineikonic mode that captures their interaction; however, when we started with Transana (and with the film *Skin*), we only had one text box and one set of time codes to represent these modes in a textual form. This led to both practical and theoretical

analytic problems.   Practically, trying to mark all four modes in a single transcript interface was difficult and resulted in a cluttered and sometimes indecipherable landscape of entwined conceptual meanings and time codes required for different purposes. While Transana afforded the capacity to link video data directly to textual transcripts and ORION encourages building multiple transcripts as means for multiple analysts to interact asynchronously around data analysis (Goldman, 2007), we were interested in using both individual transcripts and their interactions as a core mechanism for our analysis of the youth-produced videos. This need emerged from our theoretical stance on the multimodal nature of film production and our understanding of the need to attend to the kineikonic mode, the construction of meaning within individual modes and in the interaction among modes and the new meaning that is created as a result (Burn & Durran, 2006; Burn & Parker, 2003). A single transcript (even multiple transcripts generated by different analysts) could not accurately represent the interactions between modes and we could not understand these interactions in a principled way.

To resolve this methodological dilemma, we developed Transana 2.3, which added support for multiple simultaneous transcripts for a single media file. In this version, we can have up to five separate transcripts, one for each filmic element; the data, time codes, and any other notes in each transcript became specific to its respective filmic element. Using multi-transcript Transana, we are able to see how the youth producers use the four filmic elements to make direct and indirect meaning in their films. While several tools for video analysis allow for the use of clips as analytic units, in Transana 2.3 clip-making decisions can be based on any combination of transcript choices, varying from one to all available transcripts. Some young filmmakers use one of the filmic elements to great effect in communicating issues of identity. The following are a small sample of these films used to introduce the tool and multimodal methodology in action.

Identity representation in film.   Since we are interested in how young people learn to make films about the stories of our lives, the content focus of the films we are analyzing is personal identity. Research on what youth learn from their participation in YMAOs indicates that these are learning environments for multimodal production that involve identity construction (Daniels, Little, Reynolds, & Sullivan, 2006). Willett et al. (2005), for example, argue "identity" features prominently in multimodal composition: "New media production is as much about producing identities and social spaces as it is about creating media…Through different media forms young people are described as performing, defining, and exploring their identities" (p. 2). Narrative theorists have argued that these stories represent identities and that the primary mechanism we have for constructing identity is through the stories we tell (e.g. Bamberg, 2004; Bruner, 1990; Polkinghorne, 1988). Analyzing the products of a rich, complex literacy practice like digital filmmaking is a critical way to make sense of how youth engage with issues of identity through the media they create.

The use of Transana 2.3 allowed us to explore how the five primary modes for communicating meaning afford filmmakers the opportunity to construct and represent identity. We use Côté and Levine's (2002) construct of a "viable social identity," a reconciliation of the way you see yourself, the way others see you, and the way you fit into the communities to which you belong as markers of identity as a signifier. This framework draws on prior work where Halverson (2005) identified the presence of the

internal (psychological) self, the social, interactive (personal) self, the self situated in broader cultural context (social) and the merging of these dimensions of identity through a narrative analysis of written stories produced by youth participants in a theatre program. In this context, youth constructed representations of a viable social identity, "primarily through literary devices that allow the authors to play with voice and point of view" (Halverson, 2005, p. 81). We apply this method to digital videos, moving from a literary analysis to a multimodal, filmic analysis to identify signifiers of identity in youth-created media.

Our analytic methodology is similar to other research that takes video as the discourse of interest in understanding how people make meaning in a social situation (e.g. Barron, 2003; Green, Skukauskaite, Dixon, & Córdova, 2007). However, unlike many prior uses of video as a discourse of interest, we are interested in understanding how youth use the tools of video production to make meaning, rather than an analysis of their everyday interactions within the video itself. Given our interest in the ways in which the tools of film are used to signify identities, we take a top down approach to the analysis of video data as we developed our coding scheme before undertaking the analysis of the bulk of our youth-produced films (Pea & Hoffert, 2007). Barron (2007) argues that there are many ways to approach the analysis of video data; the most important issue is that analytic decisions be meaningful and well documented. We take this one step further, arguing that analytic choices must maintain the integrity of the data as fundamentally multimodal where meaning is made both within and across modes. What sets our approach apart, then, is our explicit attention to different modes of communication (filmic elements) and how these modes interact with one another as a part of the coding and analysis process.

Multimodal Methodology in Action

While the purpose of this article is to describe our methodology and companion tool for the multimodal analysis of youth-produced videos, we want to provide several illustrative examples of how our approach has allowed us to understand the relationship between youth media production and "identity work" (Willett et al., 2005). In this section, we will first describe how our multimodal method and tool reveals the role of individual modes (such as cinematography) as well as the kineikonic mode in helping one young filmmaker to understand the connections between the way she sees herself and the way others see her, Then, we will unpack how one of our case study organizations, Reel Works Teen Filmmaking, worked with its young filmmakers to use transitions as semiotic meaning making spaces for negotiating identity. While not representative of the entire corpus of films in our study, our purpose here is to demonstrate our methodological approach in action.

Understanding the use of modes in *The Mizz Perception of Roro!*

The Mizz Perception of Roro! is an autobiographical film about a young African-American woman who seeks to "tackle the misperceptions people have about tall women" (ListenUp!, n.d. a). The film can be viewed in its entirety from this link: http://listenup.org/screeningroom/index.php?view=a4dcd000c6dc5c7bb20a753f117c447f

.The film centers on her peers' initial impressions of her, how those perceptions have changed over time, and how these perceptions are connected to her own sense of herself. Overall, the film is focused on the exploration of a viable social identity (Côté & Levine, 2002), as Roro describes how she sees herself, how other people see her, how she fits into this community of teen media makers and, importantly, how these versions of self merge together. The filmmaker's use of cinematography as a tool for meaning making is especially apparent in this 10-second clip of the film. The filmmaker explores the use of extreme angles, first with a bird's eye view shot from the point of view of Roro followed immediately by an upward tilt shot from the point of view of the person passing Roro on the street. Even without sound or the visual reactions of the onscreen personalities, the filmmaker conveys her perspective of others and how others see her using exaggerated camera angles. The bird's eye view shot gives the audience the perception that they are bigger than what they are looking at. Likewise, the upward tilt conveys a feeling of smallness. The filmmaker seems to be using camera angles as a method for putting the audience first in the shoes of Roro's friend and then immediately into her own shoes. Specifically, in comparing formal filmic coding and identity coding, we found a correlation between the "how I see myself" identity code and the "bird's eye view" shot code and the "how other people see me" identity code and the "upward tilt shot" code.

When we add sound into the interpretation, the filmmaker's voiceover provides additional meaning to the shot choices. She first asks, "Why do people view me like that or…you know what I'm sayin'? What goes through people's minds when they see me?" This question is in direct response to the prior scene where an interviewee tells Roro that people think that she is "mean" and "tough" when then first meet her. The first person, bird's eye view shot gives the impression that Roro is much bigger than the person in the frame, providing explicit evidence that Roro's size may be intimidating to others. When we consider the mis-en-scene, Roro's wave to the camera in the second, upward tilt shot provides a direct counterpoint both to the interviewees' descriptions of her as mean and tough and to the prior bird's eye view shot that can appear visually intimidating. The kineikonic mode – the interaction between the cinematography and the mis-en-scene – offers a more nuanced portrait of how Roro is representing the interaction between the way she sees herself and the way others see her. Her wave directly to the camera provides a counterpoint both to the comments made in an earlier scene and to the intimidating bird's eye view shot.

Baldry and Thibault (2006) describe the introduction of new elements (or modes) into a video-based representation as a method for expressing complex meanings. Each time a new element is introduced, it acquires salience or newness as the other elements remain constant. As viewers, we attend to the new element precisely because it changes within the context of otherwise stable elements within the film. Introducing new modes layered on top of existing modes leads the audience to attend to the new element precisely because it changes within the context of otherwise stable elements within the film. In the clip described above, the filmmaker first establishes the bird's-eye-view shot before introducing the voiceover then maintains the voiceover while switching the shot to an upward tilt shot. Each time we are introduced to a new element, the extreme shot, the voiceover, the change in shot, this mode becomes the salient feature we attend to. Each transcript (and mode) represents a separate analytic lens with which we view our data, and the simultaneous display of these transcripts, combined with the way they are linked

to the video in Transana, allow us to create a tool focused on the interaction across and between transcripts and a method that displays the kineikonic mode in action.

The Importance of Transitions: Reel Works Teen Filmmaking

Reel Works Teen Filmmaking was the subject of one of our four case studies that explored how out-of-school organizations work with adolescents to produce films about the stories of their lives (Halverson et al., 2009; Halverson & Gibbons, 2010). Reel Works is a New York City-based non-profit organization founded in 2001 by two filmmakers interested in bringing the art of filmmaking to youth. Every participant enters Reel Works through The Lab, a 20-week program where 12 adolescents write, shoot, and edit short-form documentaries about the stories of their lives or issues that are important to them. The Reel Works Executive Director describes the films as either autobiographical, "or presented through an autobiographical lens" (Personal Communication, 2007). Participants are shepherded through the process by their own mentor, a professional filmmaker or editor who supports that participant at every stage. By requiring each participant to work on an individual film that tells a story about their lives, the organization places the development of individual identity at the center of the process, while at the same time supporting learning about the art of film within the constraints of short-form documentary.

First-time filmmakers at Reel Works are expected to follow a standard structure, to create a three-act documentary with transitions that link each act visually and thematically. These expectations are reflected in their pedagogy and in their discourse about the process. Over the course of their weekly lab meetings, youth watch films from previous seasons to reflect on the structural and narrative elements of these films and the relative success of these choices. In the Reel Works filmmaking process, transitions between acts are accorded special status; in these spaces filmmakers are expected to tie the film together through the introduction of their personal perspective, what the Executive Director refers to as "the I," without which the film would not be successful. Transitions are the final pieces to be produced, saved until the three acts are completed and filmmakers have a sense of their personal identity as represented in the film:

> From a storytelling perspective, no one on camera is saying that we need to make this transition from here to there, and so basically in order to change the subject we have to say something…So we structure [the film] around these spaces where this voice over is going to be. (Personal Communication, 2007)

Our multimodal methodology and tool, Transana 2.3, afforded us the ability to analyze how Reel Works filmmakers used transitions between acts to insert "the I", to explicitly represent the nexus of how they see themselves, how others see them, and how they fit into the communities to which they belong.

Hopeful Home.   In *Hopeful Home*, the filmmaker describes his experience with homelessness and the impact this experience has had on his life. While *Hopeful Home* does not follow the three-act structure as closely as other Reel Works films, the filmmaker uses the tools of film to powerfully represent a viable social identity. The story

of one young man's homelessness is told primarily through interviews with himself and his mother. The only faces shown throughout the film are the filmmaker, his mother, and one quick shot of his sister. Of the eight phases that make up this film, six focus exclusively on the filmmaker and his mother. <u>The only two phases that do not show them are two "man on the street" interviews shot with a handheld camera in black and white of others who are or have been homeless</u>. The grainy, handheld shots of life in the shelters do not show the faces of the interviewees and convey a stark contrast with the close-up shots of the filmmaker and his mother who look directly into the camera to tell their story. These shots distance the filmmaker from self-as-homeless, constructing his identity as independent from the homeless community (e.g. how I fit into the communities to which I belong; Halverson, 2005).

This is a story about how individuals in a family unit made it through the experience of being homeless. It was not about all homeless people nor was it a message to others in similar situations about how to make it through. Instead, this film is a personal look at one family's journey, their individual journey and establishes how this experience shaped the young filmmaker's viable social identity. He is the central focus of the journey but his mother's supporting role should not be downplayed, as her impact on him during their homeless situation was vital to his individual identity from this experience. This is captured in his last words in the film:

> In the end, I have to thank my mother for keeping me strong and keeping
> our whole family strong. Because she taught me that, you know like, life is
> tough and stuff happens and you always have to, you always have to keep
> your head up so you can make it through anything.

Ryan's closing monologue is followed by one final monologue from his mother, the last scene of the film. She is sitting in front of the same white wall for all of her other scenes. Her final words are very inwardly focused and do not necessarily include her family's point of view: "My very livelihood and my shelter, you know, the roof over my head, had been threatened. I had never been in that place and um, you know hey, (gives a slight shrug of her shoulders) made it through". Ryan's story begins with a black screen and an overarching statement about people's thoughts on the homeless. He was once one of those people but after his individual family experience, he knows differently. He ends the film by having the person who helped him make it through express these words herself. The choice to use his voice as well as his mother's demonstrates how his mother's view plays a large role in the construction of his viable social identity.

Journey to the Unknown.   The decision of whether and how to speak in your own film is also an explicit choice about "the I" that communicates to the audience how the director sees him or herself in relationship to the story. Filmmakers can choose from among different modes to represent themselves on film. They can appear in the film and speak in real-time through interviews or dialogue (mis-en-scene and sound), use voiceover (sound only), use video or photo footage of themselves (mis-en-scene only) or not appear at all. For example, *Journey to the Unknown* tells three interrelated stories about the impact of teen pregnancy:

My film is about three people, two in the present and one in the past. My friends Marilyn and Yasmine both became pregnant last spring. I know about teen pregnancy because my mom was a teenage when she had me. I wanted to follow my friends through their pregnancies and also to learn more about my own birth, and the circumstances surrounding it (ListenUp, n.d. c).

The filmmaker employs the Reel Works trope of using transitions to insert personal voice and identity; she marks transitions between stories of the impact of teen pregnancy with her own Quinceañera celebration. While she is always present in the shot, she does not talk directly to the camera. Rather, she couples her voiceover dialogue with b-roll footage to direct the viewer to the important visual images in the scene. In the opening phase, we see the filmmaker entering a room ballroom with a slight nervous, bashful smile on her face dressed in a pretty, bustled gown accompanied by music. She explains how she has now entered into her Mexican-American community as a grown woman through the rite of Quinceañera as she explains the ceremony through a voiceover: "Recently I turned sixteen. My family had a party for me. In Mexico it's called a Quinceañera, when a young girl comes of age."

Later in this same introductory phase of the film, as we see her dancing with a young gentleman, she tells us how she felt at the party: "It was so much fun having a party and getting dressed up. And showing everybody I'm not a little girl anymore. But I have friends who have come of age in a different way". In this voiceover, we hear the filmmaker distinguish herself from her friends, though at this point in the film, we do not know the different way they have come of age. She identifies herself as having "come of age" both through the footage of her Quinceañera coupled with her voiceover analysis of what this party meant to her development. The voiceover also propels the story forward by providing a link to the upcoming scenes in the film, which focus on a decidedly different path to womanhood. Taken together, the visual imagery and voiceover display a viable social identity – how the filmmaker sees herself (as different from her peers), how others see her (as a young woman at a coming-of-age party), and how she fits into the communities to which she belongs (as a Mexican-American but not as a teen mother).

Jewmaican. Finally, we want to return to *Jewmaican*, the film that we described at the beginning of this article, to highlight how using a multimodal methodology allows us to untangle the complex relationship between the elements of filmmaking and identity representation. Recall that the film follows the story of a young woman who struggles to reconcile her biological identity as a Jewish American and her adopted identity as the daughter of her Jamaican nanny. The filmmaker explores the tension between three identities - how she sees herself, how others see her, and how she fits into these communities (Halverson, 2005). She uses interviews with biological and adopted family members to represent how others see her, close-up shots of herself to represent how she sees herself, and b-roll visual imagery and music to represent how she fits into the communities to which she belongs.

In particular, in the transition between the first act and the second act, the filmmaker establishes her membership in the Jamaican community expanding on Jewish American identity she had represented up until this point. Figure one is a screenshot of

our analytic method in action within Transana 2.3, a very different representation from the flat transcript we presented at the beginning of this article. The transition begins with the introduction of music, Bob Marley's "Is This Love". This the first time we hear music in the film – four minutes into a seven-minute film – this new sound is particularly salient at this point in the film. Bob Marley is one of the most recognized Jamaican artists of all time, providing an audio trope for listeners to recognize the scene as explicitly Jamaican. Simultaneously, the scene shifts from medium shot interviews to footage of natural scenery from her Jamaican vacations to exemplify the differences between her home in New York and her second home and family in Jamaica. She also includes by way of visual images people, most likely members of her Jamaican family, naturally interacting with one another and with the sand and surf of the Jamaican landscape. Her use of a handheld camera is a stark juxtaposition to the tripod-mounted interviews she uses for the remainder of the film. She uses voiceover to convey her relationship with her Jamaican family, a combination of her own voice and Bob Marley's:

> Filmmaker: So along with a new mom came siblings, cousins, uncles, aunts, friends. We had fun in Jamaica, we had fun in New York…
> Bob Marley: I wanna love ya
> Filmmaker: But the fun didn't last
> Bob Marley: And treat you right
> Filmmaker: When I was eight years old
> Bob Marley: I wanna love ya
> Filmmaker: Beverly and Nikki moved to Florida.

She breaks her own voiceover to allow certain lyrics to be included in the narrative description. Her understanding of self is reflected in the combination of her voiceover, Bob Marley's lyrics, and the images of Jamaica that document her "other family". Through the explicit use of different filmic techniques throughout this phase, including the addition of a soundtrack and interweaving dialogue with lyrics in sound and switching from a tripod mounted camera to handheld footage, the filmmaker establishes her Jamaican-ness, as an additional, but essential, part of her viable social identity.

Multimodality, Video Production, and Identity

The co-development of an analytic methodology and method for working with multimodal data emerged as we sought to understand how identity is signified in youth-produced films. We adapted Transana in order to work with multiple, simultaneous transcripts. This resulted in our ability to see the kineikonic mode in action and to understand not just how youth use the individual modes of film to make meaning, but how these modes work together to produce something new. Working with these multiple transcripts allowed us to move away from the prioritization of oral production in analyzing youth representations of identity. Since *Skin* had no dialogue, we conducted our analysis with a focus on the mis-en-scene, editing, and music choices. Here again we can see how the combination of modes makes clear that youth producers are building representations of identity that reflect not just their own views of themselves but how they fit into their communities. In *The Mizz Perception of Roro* we see how

cinematographic choices, integrated with voiceover from the filmmaker communicate how she sees herself, how others see her, and how she fits into her community of peers.

Figure 1. *Screenshot of multimodal transcript for Jewmaican in Transana.*
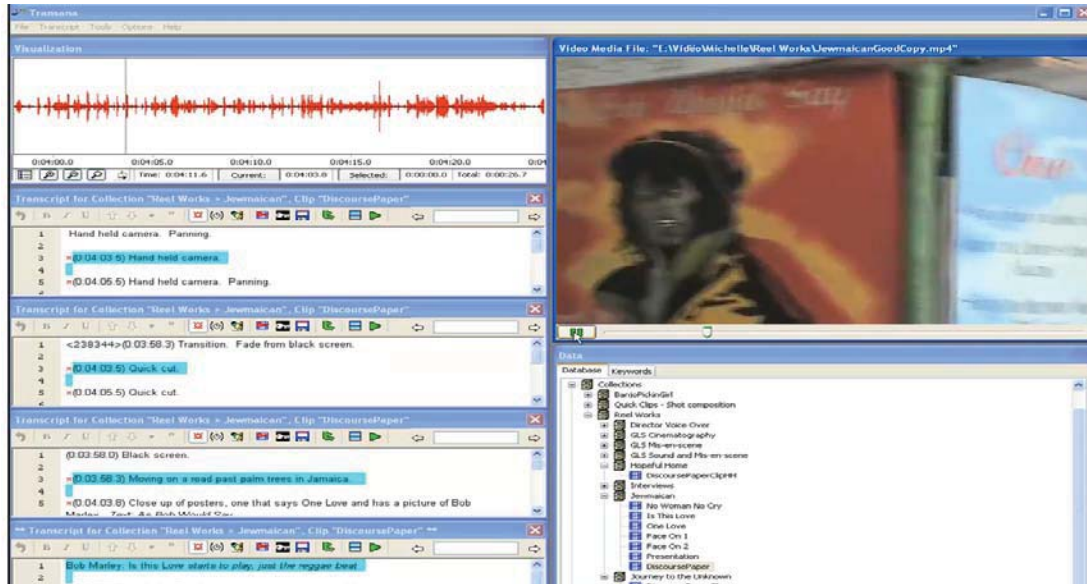
Note: The screen capture of the transition between the first and second scenes in Jewmaican helps us understand the many complex modes and meaning making that occurs during a multimodal production.

Bringing this analytic framework and tool to a larger dataset, we were able to understand how the Reel Works pedagogical approach resulted in youth producers' construction of identities. The Reel Works staff made clear to us that transitions between acts marked places in youth films where producers could insert their personal voice, their stance on the events of their lives. *Jewmaican*, *A Hopeful Home*, and *Journey to the Unknown* all demonstrate the use of transitions as spaces for youth to assert a viable social identity. Using our multimodal analytic framework and tool demonstrates how adult mentors structure opportunities for youth to engage with complex issues of identity through the creation of autobiographical films. These findings are consistent with our prior work that has demonstrated how the construction of autobiographical artistic products allows youth to construct and display a viable social identity in action (Halverson, 2005). In films, this display is most readily seen in the transition spaces between phases, where multiple modes often move separately. The relationship between these views of self is revealed in the interactions among modes, how youth producers use sound and mis-en-scene, in combination with specific editing and cinematographic choices to display the complex relationship among these views.

While it may be possible to capture multiple modes of communication in a more traditional running transcript, the visual representation of a single textual transcript does not easily afford an analysis of the interaction among modes. In fact, it likely requires the favoring of one mode, typically audio, where, "nonverbal behavior such as gesture, posture, emotional expression, and actions *might* also be described" (Barron, 2007, p. 174, emphasis added). What would we have missed in our analysis by using a single

transcript that favors audio text? Most simply, we would miss the ways in which youth filmmakers employ five different modes singularly and in combination to construct and represent identity. The choice to use both static and hand-held shots and to juxtapose them against one another, for example, is illuminated clearly when the modes are treated independently. Most significantly, perhaps, is that we would miss how the modes interact with one another to produce meaning as in the use of editing, mis-en-scene, and music in *Skin* or different types of sound in Jewmaican.

The Value of Multimodal Analytic Methodology Beyond Media Products

Research in discourse studies has broadened our understanding of literacy beyond text to embrace visual, aural and embodied modes of representation. In this work, researchers have begun to account for the multimodal nature of discourse, the understanding that "humans use many means made available in their cultures for representation precisely because these offer differing potentials, both for representation and for communication" (Kress, 2000, p. 194). These researchers have developed theoretical accounts of how people make meaning multimodally (e.g., Eggins, 1994; Kress, 2003; Kress & van Leeuwen, 2006; Lemke, 2002), and are developing methodological tools to reflect these theoretical accounts (e.g. Goldman, 2007; Goldman et al., 2007; Pea et al., 2004). For a long time, technology constrained how researchers were able to capture meaning making; qualitative data has traditionally consisted of text and audio-based accounts of human interaction. The ubiquity of video recording equipment and the recent trend in educational research to use this equipment as a primary tool for data collection (Goldman et al., 2007) has allowed us to record and store multimodal acts, comprised of words, actions, tools, and composites of technological materials. Advances in technology have led qualitative researchers to use video data as multimodal text for analysis, though we have not yet fully realized a vision for how to address the multimodality of these data. As Jay Lemke (2007) describes, "…we cannot understand the epistemology of video as representation unless we also understand the processes by which we make meaning with video when we experience it" (p. 40).

The prospect of multimodal approaches to analysis provides an exciting direction for researchers that seek to understand how people make meaning in context. For example, a teacher standing in front of a classroom trying to explain fractions may use a variety of tools in her explanation including words, pictures, physical objects, and even her physical gestures (Alibali & Nathan, 2007). Researchers may very well videotape this teacher's lesson to capture how she engages with these multiple modes to communicate meaning to and with her students. But how does the researcher then analyze this video data? Likely s/he creates a transcript of the video, and treats the transcript text as the discourse of interest (Goldman et al., 2007; Wood & Kroger, 2000). This textual dialogue-based transcript is assumed to be the most useful tool for analysis; typically, we have assumed, "that the affordances of written verbal texts far outstrip what can be offered by or offered in conjunction with other modalities" (Hull & Nelson, 2005, p. 227). However, a transcript that prioritizes oral production or is only able to capture words and/or actions separately forces the researcher to collapse multimodal data and focus on one mode of analysis. In other words, while a multimodal framework is a well-accepted theoretical frame for the collection of data around complex systems, we are just

beginning to understand how to use new technologies to capture multimodal meaning-making and how to employ that multimodal framework in data analysis (Goldman et al., 2007). Specifically, we lose the capacity to understand how different modes interact with one another, and how new meanings are made possible through these interactions.

While our work has focused on identity as signified in youth films, the analytic framework and tool we have developed is not limited to identity as an analytic construct. In the broad sample of youth films we reviewed beyond our case study work with YMAOs, other concepts being signified included: concerns affecting teens more generally such as underage drinking, teen pregnancy, and bullying, as well as civic issues such as environmental protection and adequate public school funding. There were also fictional films in our sample as well, signifying a variety of concepts including superheroes and friendship. This analytic methodology and the importance of aligning data analysis method and tool around multimodality could also be used to understand how youth represent non-identity related topics about which they feel passionate.

More broadly, research agendas in educational psychology, the learning sciences, and narrative theory could benefit from the development of analytic methods and tools focused on multimodality and meaning making. The notion of a multi-transcript feature of a video analysis tool is applicable across video analysis tools used for educational research (Pea & Hoffert, 2007). Other video analytic tools such as DIVER and ORION afford multiple analytic takes on the same video footage. Our emphasis on the use of multiple modes to convey meaning, and particularly the use of the kineikonic mode complements the growing body of analytic work that attends to the complexities of video data in educational research. Since we are interested in both what is in the video and how the video is made, the reason for an in-depth analysis of the interaction of modes is clear. Attending to how producers make meaning with the digital video medium is a potentially useful contribution to the analysis of video in the learning sciences.

In addition to similar video tools, there are also specific examples of the utility of multimodal analyses of video data including Noice and Noice's (2007) research on the role of movement in conveying meaning during a theatrical performance. While the video itself is not multimodal, performance as a representational act means that actors use modes independently and in combination (the kineikonic mode) to purposefully convey meaning just like filmic modes. This is particularly true when actors make non-literal choices with their movements and gestures in an effort to, "intentionally disambiguate the text by their specific, concrete interpretations" (Noice & Noice, 2007, p. 83).

In several research projects, the analytic alignment that would allow for the consideration of the interaction among modes is already in place. Research on the role of gesture in meaning making has extended from sign language communities (Kress, 2000), to math and science teaching and learning (Alibali & Nathan, 2007; Singer et al., 2008). Most of this research relies on video data to understand the role of gesture and its relationship to other modes of communication including speech, text, and the use of interactive technologies. Woods, Nathan, and Bieda (2007) constructed an analytic scheme for studying gesture in the teaching of mathematics based on single textual transcripts, segmented based on gesture, with verbal information super-imposed as a second layer. Separating gesture and verbal layers into separate transcripts allows for both independent analysis of these modes of instruction as well as exploration of how they are (and are not) interrelated. This research study is currently using Transana 2.3 to

create simultaneous, interrelated transcripts of classroom data that allow for exploration of the kineikonic mode in meaning making (Halverson, 2010; Woods & Dempster, 2011)

Ziegler and Woods (2008) discuss the utility of this technology in studying language use and development in pluralingual environments such as Luxembourg's schools. Children starting school in Luxembourg are thrust into a situation where the other children around them speak a variety of different languages in their homes. Portuguese, French, and German are common, but other languages such as Luxembourgish and Italian may also be present. As the children learn to communicate with each other, some very interesting things occur grammatically and linguistically. To further complicate the analysis, some of the researchers on this project do not speak all of the languages present in the classroom. In this situation, transcripts in different languages, mixing original speech with clearly labeled translation as appropriate, makes the classroom discourse more easily accessible to everyone on the research team.

Sheon (personal communication, 2008) describes a model for using this tool to enhance the study and practice of training counselors using Interpersonal Process Recall (IPR). In IPR, video (or audio only if necessary) of a counseling session is collected, and subsequent reflections on the session from the client and independently for the counselor are then captured. The original session talk constitutes the first transcript, and the client and counselor commentaries, linked to the segments of the session they are discussing, form two additional transcripts. This allows counselors in training, their supervisors, and researchers to examine the original counseling session in detail and to gain access to additional insight from the participants' commentaries, which sometimes agree and sometimes diverge sharply.

Some Final Words about the Complexity of Multimodality as Theory and Method

Just as our method and tool was two years in-the-making, we have been working on this article for three years. We say this not to complain about the challenges of collaborative research and writing, but rather to highlight how difficult it has been to wrestle to the ground the relationship between theory and method. While we began with the broad understanding of multimodality as a theory for how young people create representations of self in film, we quickly moved toward multimodality as a method for breaking apart the young filmmakers' representational choices. This move meant we had to create a way of seeing which did not reduce the analytic complexity to already existing methods for interpreting others' discourse. It also meant we needed a tool that could account for analytic complexity. We now believe that our method and tool can be used to interpret other multimodal acts beyond youth film production and beyond the construction of identity using film. We are anxious to see how other researchers who are committed to multimodality as a theoretical framework for understanding human interaction take up the methodology and tools we have offered here.

References

Alibali, M., & Nathan, M. (2007). Teachers' gestures as a means of scaffolding students' understanding: Evidence from an early algebra lesson. In, R. Goldman, R. Pea, B.

Barron, & S. Derry (Eds.), *Video research in the learning sciences* (pp. 349-366). Mahwah, NJ: Lawrence Erlbaum.

Bamberg, M. (2004). Narrative discourse and identities. In J. C. Meister, T. Kindt, W. Schernus, & M. Stein (Eds.), *Narratology beyond literary criticism* (pp. 213-237). New York, NY: Walter de Gruyter.

Baldry, A. P. (2006). The role of multimodal concordancers in multimodal corpus linguistics. In T. D. Royce & W. L. Bowcher (Eds.), *New directions in the analysis of multimodal discourse* (pp. 173-193). Mahwah, NJ: Lawrence Erlbaum.

Baldry, A., & Thibault, P. J. (2006). *Multimodal transcription and text analysis: A multimedia toolkit and coursebook*. London: Equinox Publishing.

Barron, B. (2003). When smart groups fail. *The Journal of the Learning Sciences, 12*(3), 307-259.

Barron, B. (2007). Video as a tool to advance understanding of learning and development in peer, family, and other informal learning contexts. In, R. Goldman, R. Pea, B. Barron, & S. Derry (Eds.), *Video research in the learning sciences* (pp. 159-188). Mahwah, NJ: Lawrence Erlbaum.

Bordwell, D., & Thompson, K. (2004). *Film art: An introduction* (7th ed.). New York, NY: McGraw-Hill.

Burn A. & Durran, J. (2006). Digital anatomies: Analysis as production in media education. In D. Buckingham & R. Willett (Eds.), *Digital generations*. Mahwah, NJ: Lawrence Erlbaum.

Burn, A., & Parker, D. (2003). *Analysing media texts*. London: Continuum.

Bruner, J. (1990). *Acts of meaning*. Cambridge, MA: Harvard University Press.

Côté, J. E., & Levine, C. G. (2002). *Identity formation, agency, and culture: A social psychological synthesis*. Mahwah, NJ: Lawrence Erlbaum Associates.

Crotty, M. (2003) *The foundations of social research,* (2nd ed.). Thousand Oaks, CA: Sage Publications.

Curwood, J. S. & Gibbons, D. (2009). "Just like I have felt": Multimodal counternarratives in youth-produced digital media. *International Journal of Learning and Media, 1*(4), 59-77

Daniels, S., Little, P., Reynolds, L., & Sullivan, A. (2006). Where visual literacy and identity meet: Adolescents define themselves through participation in a university video and art enrichment program. *Wisdom in Education*. Retrieved from http://www.wisdomineducation.org/journal/fall_2006/where_visual_literacy_and_identity_meet_adolescents_define_themselves_through_participation_in_a_uni.php

Eggins, S. (1994). *An introduction to systemic functional linguistics,* (2nd ed.). London: Continuum International Publishing Group.

Gee, J. P. (2003). *What video games have to teach us about learning and literacy*. New York, NY: Palgrave/Macmillan.

Goldman, R. (2007). Representations and the perspectivity framework: Epistemology, ethnography, evaluation, and ethics. In R. Goldman, R. Pea, B. Barron, & S. Derry (Eds.), *Video research in the learning sciences* (pp. 3-38). Mahwah, NJ: Lawrence Erlbaum.

Goldman, R., Crosby, M., Swan, K., & Shea, P. (2005). Qualitative and quisitive research methods for describing online learning. In S. R. Hiltz & R. Goldman (Eds.),

       Learning together online: Research on asynchronous learning networks (pp. 103-120). Mahwah, NJ: Lawrence Erlbaum.

Goldman, R., Pea, R., Barron, B., & Derry, S. J. (2007). *Video research in the learning sciences*. Mahwah, NJ: Lawrence Erlbaum.

Green, J., Skukauskaite, A., Dixon, C., & Córdova, R. (2007). Epistemological issues in the analysis of video records: Interactional ethnography as a logic of inquiry. In, R. Goldman, R. Pea, B. Barron, & S. Derry (Eds.), *Video research in the learning sciences* (pp. 115-132). Mahwah, NJ: Lawrence Erlbaum.

Halverson, E. R. (2005). InsideOut: Facilitating gay youth identity development through a performance-based youth organization. *Identity: An International Journal of Theory and Research, 5*(1), 67-90.

Halverson, E. R. (2010). Film as identity exploration: A multimodal analysis of youth-produced films. *Teachers College Record, 112*(9), 2352-2378.

Halverson, E. R. (in press). Digital art-making as a representational process. *The Journal of the Learning Sciences.*

Halverson, E. R., & Gibbons, D. (2010). "Key moments" as pedagogical windows into the video production process. *Journal of computing in teacher education, 26*(2), 69-74.

Halverson, E., Lowenhaupt, R., Gibbons, D., & Bass, M. (2009). Conceptualizing identity in youth media arts organizations: A comparative case study. *Journal of E-Learning, 6*(1), 23-42.

Hull, G. A., & Katz, M. (2006). Crafting an agentive self: Case studies of digital storytelling. *Research in the Teaching of English, 41*(1), 43-81.

Hull, G. A., & Nelson, M. E. (2005). Locating the semiotic power of multimodality. *Written Communication, 22*(2), 224-261.

Jenkins, H., Purushotma, R., Clinton, K., Weigel, M., & Robison, A. (2007). *Confronting the challenges of participatory culture: Media education for the 21st century. Building the field of digital media and learning*. Retrieved from http://www.digitallearning.macfound.org

Kola, G., Kosar, T., & Livny, M. (2004). A fully automated fault-tolerant system for distributed video processing and off-site replication. In *Proceedings of the 14th international Workshop on Network and Operating Systems Support for Digital Audio and Video* (Cork, Ireland, June 16 - 18, 2004, pp. 122-126). New York, NY: NOSSDAV '04. ACM.

Kress, G. (2000). Multimodality. In B. Cope & M. Kalantzis (Eds.), *Multiliteracies: Literacy learning and the design of social futures* (pp. 182-202). London: Routledge.

Kress, G. (2003). *Literacy in the new media age*. New York, NY: Routledge.

Kress, G., & van Leeuwen (2006). *Reading images: The grammar of visual design.* (2nded.) New York, NY: Routledge.

Lemke, J. (1998). Multiplying meaning: Visual and verbal semiotics in scientific text. In J. R. Martin & R. Veel (Eds.), *Reading science* (pp. 87-113). London: Routledge.

Lemke, J. (2002). Travels in hypermodality. *Visual Communication, 1*(3), 299-325.

Lemke, J. (2007). Video epistemology in-and-outside the box: Traversing attentional spaces. In, R. Goldman, R. Pea, B. Barron, & S. Derry (Eds.), *Video research in the learning sciences* (pp. 39-52). Mahwah, NJ: Lawrence Erlbaum.

Lemke, J. L. (2010). Transmedia traversals: Marketing meaning and identity. In A. Baldry & E. Montagna (Eds.), *Interdisciplinary perspectives on multimodality: Theory and practice*. Proceedings of the Third International Conference on Multimodality. Campobasso: Palladino.

LeVine, P., & Scollon, R. (Eds.). (2004). *Discourse and technology: Multimodal discourse analysis*. Washington, DC: Georgetown University Press.

ListenUp! (n.d. a). *The Mizz Perception of Roro!* Retrieved from http://listenup.org/screeningroom/index.php?view=a4dcd000c6dc5c7bb20a753f1 17c447f

ListenUp! (n.d. b). *SKIN*. Retrieved from http://listenup.org/screeningroom/index.php?view=c76a8891ddf0739d8e83ce66ec bf2d3a

ListenUp! (n.d. c). *Journey to the Unknown*. Retrieved from http://listenup.org/screeningroom/index.php?view=54fab213eaf7eaf300208d2553 490e09

Manovich, L. (2002). *The language on new media*. Cambridge, MA: MIT Press.

Mavrou, K., Douglas, G., & Lewis, A. (2007). The use of Transana as a video analysis tool in researching computer-based collaborative learning in inclusive classrooms in Cyprus. *International Journal of Research & Method in Education, 30*(2), 163-178.

Mitchell, W. J. T. (1994). *Picture theory: Essays on verbal and visual representation*. Chicago, IL: University of Chicago Press.

Mondada, L. (2006). Video recording as the reflexive preservation and configuration of phenomenal features for analysis. In H. Knoblauch, J. Raab, H. G. Soeffner, & B. Schnettler (Eds.), *Video analysis: Methodology and methods* (pp. 51-68). Berlin: Peter Lang.

Nelson, M. E., Hull, G., & Roche-Smith, J. (2008). Challenges of multimedia self-presentation: Taking, and mistaking, the show on the road. *Written Communication, 25*(4), 415-440.

Noice H., & Noice, T. (2007). The non-literal enactment effect: Filling in the blanks. *Discourse Processes, 44*(2), 73-89.

Norris, S. (2004). Multimodal discourse analysis: A conceptual framework. In P. LeVine & R. Scollon (Eds.), *Discourse and technology: Multimodal discourse analysis* (pp. 101-115). Washington, DC: Georgetown University Press.

NVivo8. (2007). Video, audio and documents. Have you ever wished you could work with all of your research material in the same way and in one software program? *QSR International Newsletter, 32*. Retrieved from http://www.qsrinternational.com/FileResourceHandler.ashx/Newsletters/Docume ntFile/18/NSight_Edition_32_July_2007.pdf

Pea, R., & Hoffert, E. (2007). Video workflow in the learning sciences: Prospects of emerging technologies for augmenting work practices. In, R. Goldman, R. Pea, B. Barron, & S. Derry (Eds.), *Video research in the learning sciences* (pp. 427-460). Mahwah, NJ: Lawrence Erlbaum.

Pea, R., Mills, M., Rosen, J., Dauber, K., Effelsberg, W., & Hoffert, E. (2004). The DIVER™ project: Interactive digital video repurposing. *IEEE multimedia, 11*(1), 54-61.

Polkinghorne, D. (1988). *Narrative knowing and the human sciences*. Albany, NY: State University of New York Press.

Probst, Y. C., DeAgnoli, K., Batterham, M. & Tapsell, L. C. (2009). Video recorded participant behaviours: the association between food choices and observed behaviours from a web-based diet history interview. *Journal of Human Nutrition and Dietetics, 21*(1), 21-28.

Reel Works Teen Filmmaking. (n.d.). Jewmaican. Retrieved from http://www.reelworks.org/watch.php.

Singer, M., Radinsky, J., & Goldman, S. R. (2008). The role of gesture in meaning construction. *Discourse Processes, 45*(4), 365-386.

Stake, R. E. (2000). Case studies. In, N. K. Denzin & Y. S. Guba (Eds.), *Handbook of qualitative research* (2nd ed., pp. 435-454). Thousand Oaks, CA: Sage Publications.

Wikipedia. (n.d.). *Feeling good*. Retrieved from http://en.wikipedia.org/wiki/Feeling_Good.

Willett, R., Burn, A., & Buckingham, D. (2005). New media, production practices, learning spaces. *Education, Communication, and Information, 5*(1), 1-3.

Wood, L. A., & Kroger, R. O. (2000). *Doing discourse analysis: Methods for studying action in talk and text*. Thousand Oaks, CA: Sage Publications.

Woods, D., & Dempster, P. (2011). Tales from the bleeding edge: The qualitative analysis of complex video data using Transana. *Forum: Qualitative Research, 12*(1). Retrieved from http:// www.qualitative-research.net/index.php/fqs/article/view/1516.

Woods, D., Nathan, M., & Bieda, K. (2007). *Data session on Transana use for qualitative and quantitative gesture analysis.* Paper presented at the International Society for Gesture Studies Conference, Evanston, IL: Northwestern University.

Zahn, C., Hesse, F., Finke, M., Pea, R., Mills, M., & Rosen, J. (2005). Advanced video technologies to support collaborative learning in school education and beyond. In D. Suthers & T. Koschmann (Eds.), *Computer Support for Collaborative Learning Proceedings of the 2005 conference on Computer support for collaborative learning: learning 2005: the next 10 years!* Mawah, NJ: Erlbaum Associates.

Ziegler, G., & Woods, D., (2008). Applying multiple analytic layers in the analysis of video using Transana. *AnaLogiQual 2008*, CRP Henri Tudor, Luxembourg.

Appendix A
List of Coding Terms

Mis-en-scene
    Subject-related
        Facial expressions
        Gestures & body movements
        Clothing & makeup choices
    Setting
    Scripted features
    Style

Editing
    Transitions
        Hard cut
        Fade
        Dissolve
    Flashback
    Flashforward
    Special effects
        Freeze frame
        Reverse motion
Cinematography
    Lighting
    Focus
    Framing & composition
    Angle
        High (bird's eye view)
        Low (upward tilt)
        Dutch angle (diagonal tilt)
    Shot types
        Long
        Medium
        Close-up
        Static
        Zoom
        Eye-level
        POV
    Camera movement
        Pan
        Tilt
        Dolly (tracking)
        Handheld
        Steadycam
    Duration of image
        Long take

Sound
    Dialogue
    Sound effects
    Music
    Diegetic
    Non-diegetic
    Internal diegetic

Note. Adapted from Bordwell, D. & Thompson, K. (2004). Film Art: An Introduction. (7th Ed.) New York: McGraw-Hill.Figure 1. Multi-transcript Transana Screen Capture of Jewmaican

Author Note

Dr. Erica Rosenfeld Halverson is an Assistant Professor of Curriculum & Instruction at the University of Wisconsin-Madison. Erica's research focuses on how young people learn to make art about the stories of their lives and the role this process plays in identity development and literacy learning. Erica's work has been published in a wide variety of outlets including *Journal of the Learning Sciences, Teachers College Record,* and *Journal of Adolescent Research* and was the 2010 recipient of the Jan Hawkins Award for Early Career Contributions to Humanistic Research and Scholarship in Learning Technologies. Erica is the co-founder of *Barrel of Monkeys,* one of Chicago's premiere arts education theatre ensembles. Correspondence regarding this article can be addressed to Dr. Erica Rosenfeld Halverson at The University of Wisconsin-Madison, 556B Teacher Education, 225 N. Mills St., Madison, WI 53706; Phone: 608.262.5210; Fax: 608.262.0843 and E-mail: erhalverson@education.wisc.edu

Michelle B. Bass is a doctoral student completing her PhD in the Department of Educational Psychology at the University of Wisconsin – Madison. The title of her dissertation is, "Exploring the representations and remixes of underrepresented students' identities during their transition to college." Her research interests include: adolescent identity development, the transition to college, first-generation college students, and digital literacy.

David Woods is a Researcher at the Wisconsin Center for Education Research at the University of Wisconin – Madison. His primary interest is in facilitating the analytic work of researchers, particularly those with challenging data or innovative analytic approaches. He is the lead developer of Transana, software for the transcription and qualitative analysis of video and audio data.

Article Citation

Halverson, E. R., Bass, M., & Woods, D. (2012). The process of creation: A novel methodology for analyzing multimodal data. *The Qualitative Report*, *17*(Art. 21), 1-27. Retrieved from http://www.nova.edu/ssss/QR/QR17/halverson.pdf