Engaging the Senses in Qualitative Research via Multimodal Coding: Triangulating Transcript, Audio, and Video Data in a Study With Sexual and Gender Minority Youth

Shelley L. Craig¹, Lauren B. McInroy², Ami Goulden¹, and Andrew D. Eaton³

Abstract
The variety of formats in which qualitative data may be collected have been explored within the methodological literature. Yet, the multiple options for coding these data formats have not been comprehensively detailed. While transcript analysis is widely used across disciplines, it may have limitations—particularly for research involving marginalized populations. This paper presents a multimodal coding approach as a methodological innovation for triangulating three data formats (transcript, audio, and video), detailed through the application of this analytic approach during a qualitative study exploring media engagement with sexual and gender minority youth (SGMY). Nineteen semi-structured interviews with SGMY were filmed and transcribed. Nine independent coders then utilized the innovative multimodal approach to code the three data formats using a constructivist grounded theory framework. Some codes were similar across modalities, such as those related to safety issues and finding identity and community through media. Others differed between modalities, such as those related to participant affect, perceived contradictions, discrepancies between verbal statements and body language, level of comfort and engagement, and distress when discussing traumatic experiences. Video coding captured the broadest range of emotions and experiences from marginalized youth, while transcripts provided the most straightforward form of data for coding. Multimodal coding may be applicable across qualitative approaches to enrich analyses and account for potential biases, thereby enhancing analytical lenses in qualitative inquiry. Methodological strategies for coding and integrating data types are discussed.

Keywords
methods in qualitative inquiry, grounded theory, constructivist GT, qualitative evaluation, social justice, multimodal coding, data types

Introduction
Qualitative researchers have explored numerous formats for data collection and their potential for combination; such as field notes, individual interviews (in-person or online), and group recordings (Archibald et al., 2019; Tessier, 2012). However, integrating coding of multiple data formats has received less attention. As video recording technologies became increasingly accessible, opportunities for research innovation emerged to incorporate multi-faceted aspects of interviews into the analytical process. In particular, contemporary computer-assisted qualitative data analysis (CAQDAS) programs support multimodal coding (Gibbs et al., 2002). Defined as synchronous coding of text, audio, and/or video data—multimodal coding is used to understand dynamics, emotions, and emphases across data formats. It may also incorporate other sources of data such as rich text, diagrams, and images (Gibbs et al., 2002). Yet despite its potential for enriching analyses, and emerging considerations of how CAQDAS-facilitated coding can be

¹ Factor-Inwentash Faculty of Social Work, University of Toronto, Ontario, Canada
² College of Social Work, The Ohio State University, Columbus, OH, USA
³ Faculty of Social Work, University of Regina–Saskatoon Campus, Toronto, Ontario, Canada

Corresponding Author:
Shelley L. Craig, Factor-Inwentash Faculty of Social Work, University of Toronto, 246 Bloor Street West, Toronto, Ontario, Canada MSS 1V4.
Email: shelley.craig@utoronto.ca
integrated with more traditional coding methods (e.g., paper-based, whiteboards), a notable literature gap exists examining multimodal coding.

This gap is particularly notable regarding differences in the modes of coding three primary data formats (i.e., transcription, audio, video), any advantages or disadvantages for using each mode, and the potential benefits of using a combination of coding modes (Maher et al., 2018). This paper introduces a multimodal coding approach—integrating coding of text, audio, and video data—followed by an exploration of its utility through a constructivist grounded theory analysis of semi-structured interviews with sexual and gender minority youth (SGMY) about their engagement with offline and online media. SGMY refers to young people who identify as lesbian, gay, bisexual, transgender, and a variety of other minority sexual and gender identities.

Background

Multimodal Analysis

An expansion of CAQDAS software in recent years has led to enhanced opportunities to integrate different types of qualitative data and modes of coding. The most common data integration approach is to synchronize transcripts with their corresponding audio and video files to create a multimodal representation (Silver & Patashnick, 2011). Arguably, a multimodal representation of data produces a more holistic analysis than textual analysis alone (Markle et al., 2011; Silver & Patashnick, 2011), enhancing rigor and lending credibility to the conclusions drawn from qualitative research (Pink, 2013). The key consideration, which this article addresses, is exploring the added value of a more holistic analysis and considering when a multimodal approach may be well suited. Additionally, annotations, memos, and codes can be synchronized to the multimodal transcript—facilitating further rigor by incorporating reflexivity. Reflexivity is typically considered only with the researchers’ positionalities; however, visual technologies may enable greater participant reflexivity. For example, child visual reflexivity is a video-facilitated method of enhancing an understanding of how children create their perspectives (Chawla-Duggan et al., 2020). Comparatively, minimal literature exists related to data integration and production of a multimodal transcript, with academics taking for granted transcript objectivity (Bezemer & Mavers, 2011; Davidson, 2009).

Multimodal coding approaches are not limited to the synchronization of text, audio, and video data. Approaches are varied and creative depending on the researcher’s resources and perspective. For example, some have combined linguistic inquiry and word counting software with thematic analysis, creating a mix of qualitative and quantitative approaches (Firmin et al., 2017). With advances in coding software and visual technologies, there have also been increasingly flexible forms of transcripts or merging of different types of data (e.g., photos and text), often called “transvisuals” (Bezemer & Mavers, 2011, p. 192). When considering multimodal research, it is important to keep in mind that perfect translation from one mode to another (e.g., visuals to text, text to sound) is impossible, which is why some transcripts try to account for this in other ways (e.g., font formatting). The advantages and disadvantages of coding each of the data formats mentioned above are summarized in Figure 1 and reviewed in detail below.

Transcript Analysis

Interview recording and transcribing (referring to converting an interview into text) became widely accessible in the 1970s and has developed into the foremost approach to formatting data collected from interviews and focus groups. Widespread adoption of transcript analysis across qualitative research traditions is attributed to its ease of use, as well as common perceptions that the conversion of data from recordings to verbatim transcripts permits participants’ exact and detailed statements to be analyzed (versus researchers’ notes or brief open survey responses; Bezemer & Mavers, 2011). Verbatim transcripts may increase the depth of data available and the rigor of the research process (Creswell, 2007; Evers, 2011; Loubere, 2017). Many researchers continue to consider transcript analysis a means to accurately and comprehensively understand the meaning of interview and focus group data (Mishler, 1986; Ross, 2010; Sutton & Austin, 2015).

While employing verbatim transcription as the basis for analysis is usual practice in qualitative research, it is not without limitations (Lapadat, 2000). First, the volume of data and level of detail in an in-depth interview transcript produces can be overwhelming and contribute to a feeling of “drowning” in data; especially if there are many interviews, or if the interviews are lengthy (Evers, 2011, p. 3). The influx of data can interfere with a fulsome summary of the results and the study’s rigor. For example, data fatigue may cause inconsistencies between researchers during analysis (White et al., 2012). Large amounts of data may also shift the focus of analysis from meaning to quantity, leading to an inadequate analysis (Seidel, 1991). As all written transcription of recordings is reductionist to some degree (i.e., it is impossible to fully translate audio-visual dynamics into a textual document), an overwhelming amount of text data may also miss important audio and visual cues (Silver & Patashnick, 2011). Second, transcribing interviews often requires a significant investment of resources (e.g., time, staff and students to assist with transcribing, funding for a transcription service). Transcribing interviews is commonly completed outside the research team. Issues have been raised with this practice, such as discrepancies between transcribers in terms of the level of detail (White et al., 2012). Tilley and Powick (2002) studied the experiences of eight transcribers who were hired to complete transcription work on a contractual basis. The authors were particularly interested in how the transcribers, external from the research team, influenced the transcripts and the consequences in the data analysis stage. The transcribers reported several challenges and barriers during the study. They identified issues related to their lack of familiarity with the language and culture connected to the research topic,
felt pressure to tidy up “the messiness” of conversation, and lack of direction from the research team about how the transcripts should be completed (Tilley & Powick, 2002, p. 300). The findings suggest that the approach to transcription is critical to the process of data analysis, and particular elements (e.g., transcribers, transcript production) should be considered during the early stages of research design.

Further complicating matters, multiple transcribing approaches exist when working with interviews. A pragmatic transcript is the most commonly produced type, as its flexibility is sensitive to the resources available to the researcher (Evers, 2011). In a pragmatic transcript, the interview dialogue is transcribed verbatim from the recording and no attempts are made to neutralize the loss of multidimensional elements of the interview, such as the participant’s speed, pace, intonation, song, hesitation, verbal utterances (Gibbs, 2010), interview context, or background noise (Evers, 2011). A Jeffersonian transcript is similar, but also includes symbols to represent sound, pace, intonation, and interaction in the conversation (Evers, 2011). The Jeffersonian transcript is perceived as the most intensive transcribing approach, as it requires significant time due to the level of detail required (Evers, 2011). Lastly, a gisted transcript is less detailed than both the pragmatic and Jeffersonian transcripts (Evers, 2011). It does not include a verbatim interview text, but a combination of multiple summaries that capture the interview.

Additionally, information regarding the decision-making around transcription—including the particular transcribing approach used—are generally absent from publications, though there have been calls for added transparency (Davidson, 2009; Skukauskaite, 2012; Tilley & Powick, 2002). Transcribing approaches have become increasingly flexible over time and are now recognized by various disciplines, analytical purposes, and epistemologies. For example, the positivist paradigm often frames interview transcripts as an objective reflection of the interview or research activity. In contrast, the constructivist paradigm considers interview transcripts as a socially constructed reflection of reality formed by external and internal processes such as the researchers’ stance and participant context (Cupchik, 2001). Consequently, constructivist scholars have asserted problems with the singular reliance on transcription analysis; even extending their criticism to more traditional grounded theory approaches (Bezemer & Mavers, 2011). We approach this work from a constructivist worldview. Specifically, a constructivist approach to grounded theory (Charmaz, 2014), which encourages multiple data sources that can then be coded to construct an understanding of participant experiences.
Some theoretical approaches (such as post-positivism) aim to create reliable coding schemes to address trustworthiness or reliability in qualitative research. However, they typically do not focus on transcribed semi-structured interviews. Most coding schemes focus on other types of data collection methods such as field notes, documents, and ethnographies (Campbell et al., 2013). While our constructivist approach aims more for “... abstract understandings that theorize relationships between concepts” (Rieger, 2019, p. 228), it is important to consider key components in other approaches so that this multimodal coding framework may be of broad benefit. One of these components is intercoder or Interrater reliability, which assesses the extent to which two or more coders are selecting the same code for the same concept during data analysis (Krippendorff, 2004). Intercoder reliability aims to make the level of agreement among multiple coders transparent and to demonstrate different interpretations of the same data (Krippendorff, 2004). The importance of intercoder reliability may vary based on the approach, method, and researchers’ positionality (McDonald et al., 2019). For example, intercoder reliability may be less useful when coding teams share many characteristics (e.g., personal and professional backgrounds) and may interpret the data in more similar ways (McDonald et al., 2019). High levels of intercoder reliability become more challenging in less structured or standardized interviews (Campbell et al., 2013). Published studies that use interview data rarely discuss if intercoder reliability, or reliability in general, was assessed (Campbell et al., 2013; McDonald et al., 2019).

Semi-structured interviews tend to produce longer transcripts than more structured and close-ended questionnaires because participants are encouraged to expand upon tangents. In effect, each interview goes in its own direction, at least partially. More structured or close-ended questionnaires typically do not need extensive coding. An increase in the amount of text and the diversity of concepts between interviews often leads to multiple codes being necessary for one section of the text, which can be a barrier to consistency across multiple coders.

Audio and Video Analysis

In the 1990s, there was an increase in the availability of digital video recording technologies and CAQDAS software (e.g., ATLAS.ti, MAXQDA, NVIVO, Dedoose). As a result, alternatives to traditional coding arose; including coding audio and video data segments directly (Bassett, 2004; Bezemer & Mavers, 2011; Evers, 2011). Visual digital technologies (e.g., video cameras, smartphones, computers that can create and display video; Chawla-Duggan et al., 2020), are experiencing a time of sustained growth in research (Bezemer & Mavers, 2011). Yet it is critical for researchers to better understand the implications of these technologies to generate research that is rigorous and irreparable (Pink, 2013).

The inclusion of visual methods can generate data that encourage a more thorough interpretation of the phenomena of interest. In studies with children, video allows for the deeper emergence of the participant perspective compared to text, which may not fully represent their experience. Visual technologies can illuminate the complexity that comprises a participant’s social or physical situation, or capture the dialectic process between participants and the interviewer (Chawla-Duggan et al., 2020). Coding audio and video data extends analysis beyond an account of the dialectic interview process (i.e., logical discussion of opinions or ideas) to engage more with emotions and affect expressed in the interaction (Chawla-Duggan et al., 2020). Semi-structured interviews may be particularly significant to analyze via video data, as less structure can result in unexpected areas of conversation and inquiry (Crichton & Childs, 2005). Video also allows for an additional level of analysis because of the diversity of data collected. For example, interactions, concurrent actions (Norris, 2004), and body movements (Bezemer, 2008) can be examined. Thus, many consider audio and video coding to be a meaningful complement to, or improvement over, transcript coding since it provides a more precise representation of the data as it was collected (Merriam, 1998) while retaining the richness of what was said and how it was said (Crichton & Childs, 2005). Incorporating audio and video analysis may initially add to the complexity of the data to be analyzed. However, analysis of text without audio or video risks changing or removing the context of participants’ stories (Crichton & Childs, 2005; Schnettler & Raab, 2008).

In addition to how useful audio and video research is during analysis, such data can also contribute to knowledge translation activities. For example, there are benefits to using audio and video clips in scholarly and community presentations and publications (Friend & Militello, 2015). Video research data has also been used as an online resource (e.g., university website, YouTube), and has been integrated into professional curricula to facilitate classroom learning (Friend & Militello, 2015). Thus, as a multipurpose tool, research incorporating video data collection can assist in furthering both knowledge mobilization to the community and evidence-based approaches to professional practice, potentially enabling a more democratic approach to research (Chawla-Duggan et al., 2020).

Sharing video data should be done with respect to data protection and anonymity considerations, including informed consent from participants on how and where their data will be shared and protecting the data through limits on how it can be downloaded and accessed (Eaton, 2019; McInroy, 2016). Ethical considerations of video dissemination are important to discuss upfront with participants in the initial consent process before the video is recorded. Otherwise, people may behave differently, or possibly be more reticent in fully participating compared to audio-recorded interviews (Brown, 2018). An ongoing consent process—wherein participants continue to share control over their image during the dissemination phase of the research project—can also help mitigate privacy concerns regarding video distribution (Craig et al., 2020).

Despite these benefits, there still is a lack of clarity around the analytical and technical procedures and multiple ways of
analyzing video data, as well as the ways to use CAQDAS software, continuing the trend in qualitative research (Bezemer & Mavers, 2011; Fielding & Lee, 1998; Rahman, 2016; Silver & Patashnick, 2011). Thus, the gap in literature focused on qualitative data analysis continues to widen. Several limitations to incorporating CAQDAS in qualitative data analysis also exist. For instance, there is a steep learning curve for some researchers who have limited experience using software and assigning research assistants is not always feasible (Rahman, 2016; Silver & Patashnick, 2011). Some academic institutions do not support purchasing software or lack the resources to do so (Atieno, 2009; Fielding & Lee, 1998), and software package licences can be limited—causing difficulties in collaboration across institutions (Silver & Patashnick, 2011).

**Analysis of Interviews With Marginalized Populations**

Important features of the interview are inevitably lost in the progression from the recorded interview to the transcript such as pace and intonation (Bezemer, 2008; Gibbs, 2010). The absence of such features may result in valuable data being overlooked, particularly elements essential to cross-cultural research and research with marginalized populations (Didkowsky et al., 2010; Loubere, 2017). Thus, alternative methods to transcribing interviews verbatim, specific to these research scenarios, are being developed. One method is the systematic and reflexive interviewing and reporting (SRIR) method, created within a cross-cultural context where language barriers existed between researchers and participants. Transcribing interviews verbatim reduced relevant data because the non-verbal communication was lost after the fieldwork was completed (Loubere, 2017). In the SRIR method, two researchers jointly conduct the interview, subsequently engage in reflexive dialogue, and write the interview and analysis reports together. The expansion of verbatim transcription in the context of Loubere’s (2017) study was needed as differences in language use and proficiency between participants and transcriptionists, such as local dialects, made it difficult to accurately transcribe the interviews.

Working with marginalized populations requires research methods that are sensitive to context and capture the complexity of their experiences. Interview methods promote the illumination of marginalized voices that may have been previously silenced (Bezemer & Mavers, 2011; Chawla-Duggan et al., 2020; McInroy, 2016). Several multimodal research techniques that offer an opportunity for an authentic reflection of the lives of participants who are often underrepresented in research have been proposed. One such technique—the Enhancing Audio Recorded Research (EARR) model—was developed to embed audio clips in poster, oral presentations, and manuscripts (Chandler et al., 2015). Within this multimodal technique, the importance of the participant’s voice in qualitative research was emphasized via audio-enhanced dissemination. Chandler and colleagues (2015) argued that enabling their audiences to experience the power of the data through listening to it would more fully honor the voices of participants. Implementation of the EARR model “enabled a deeper expression of the findings by revealing voice inflection, tone, and emotion that are often difficult to communicate through traditional dissemination channels” (Chandler et al., 2015, p. 4).

Another multimodal method emerging as a data collection and analysis technique specifically for research exploring resilience in marginalized youth is the integration of visual qualitative data with interviews (Didkowsky et al., 2010). In this approach, visual data includes photography and videotaping of youth participants due to the perception that researchers may have unintentional difficulty understanding and representing the unique experiences of youth using verbatim transcription. This may be partly due to the possibility of participants having limited vocabulary to discuss certain topics and experiencing difficulty communicating precisely what they mean (Didkowsky et al., 2010). Researchers may also find it challenging to appreciate the context encompassing the narratives, leading to a distorted analysis of the data.

Some studies have also challenged the typical roles and responsibilities of community members involved in research projects (i.e., peer researchers), who typically collaborate on study design and recruitment efforts, but not data analysis. Sweeney and colleagues (2013) incorporated service users as coders in a study on cognitive behavioral therapy. They found that agreement among researcher and peer analysts was high overall, yet several important differences were identified. For example, when coding for experiences the researcher identified a variety of symptoms and emotions, while the service user highlighted coping strategies. Such differing perspectives are necessary to consider in qualitative analysis, as assumptions of shared knowledge may be challenged during in-depth coding (Eaton et al., 2018). Thus, a multimodal analysis can be a way to uncover varied perceptions. For example, interpretations of emotions can be based on the format, so intercoder agreement may be better achieved by analyzing multiple formats of the same data source (Craig et al., 2020).

**Application of Multimodal Coding With Sexual and Gender Minority Youth**

A qualitative study using constructivist grounded theory was conducted with SGMY (n = 19) in Toronto, Canada. The study sought to explore how SGMY experience media offline (e.g., billboards, cable television) and online (e.g., gaming, social media), as well as the impact of such experiences on their resilience and identity development. The study was also designed to explore the utility of multimodal coding with this marginalized population. In-depth semi-structured interviews were conducted with SGMY participants (aged 18–22), ranging from 45 to 90 minutes in length, and were simultaneously audio and video recorded.

The study was compliant with a University of Toronto Health Sciences Research Ethics Board protocol (ID#26749). Participants were recruited over a 3-month period via email outreach to organizations serving SGMY in the region. Participants were eligible to participate if: (a) they identified as
Grounded theory is one of the most frequently used qualitative approaches (Bryant & Charmaz, 2007; Charmaz, 2014). It is a systematic method to analyze interviews, interactions, and contexts that are part of collected data to develop theories about specific phenomena grounded in that data. Grounded theory identifies that reality and related theories are processes that are substantiated in context, and researchers are charged with consistently identifying important constructs iteratively (Corbin & Strauss, 2015). As it aligns with the researchers’ professional and epistemological worldview, this study specifically utilized constructivist grounded theory developed by Charmaz (2000). This approach integrates participants’ experiences, perspectives, and feelings to ensure that data and analysis are produced through collaboration (Charmaz, 2014). Constructivist grounded theory recognizes that participants attribute meaning to their lives and act accordingly. Consequently, reality and action are inexorably linked (Charmaz, 2014). A particular focus is the analytical process that strives to articulate relationships of concepts in a broader theoretical or explanatory framework. Charmaz (2000) cautions that to avoid trivial analysis or unsatisfactory data researchers should be aware that their own observations may not accurately depict participants’ experiences and that participants’ assumptions may be more important than their words. Constructivist grounded theory suggests that those investigating new phenomena should remain open to new insights, while also retaining their existing knowledge.

Data Management & Analysis Framework

The data management and analysis framework designed for this study utilized the multimodal data analysis approach to produce a more holistic analysis and a better representation of the interview data. The interview data prepared for analysis in three formats: (a) transcripts using the pragmatic Jeffersonian format with embedded timecodes (Evers, 2011); (b) audio files; and (c) video files. Data analysis for all formats was undertaken using the CAQDAS program ATLAS.ti. Nine independent coders were selected to participate in data analysis, representing a range of disciplines (e.g., education, social work, psychology), education levels (e.g., undergraduate, graduate, post-graduate), racial identities (e.g., African-Canadian, South-Asian, White), and sexual identities (e.g., lesbian, gay, straight, bisexual). Most coders identified as peer researchers, aligning themselves with sexual and gender minority young people.

The coding framework for the study allowed comparisons of different coding formats. Coders one through eight were asked to code 12 interviews using data formatted only a single way (i.e., transcript data or audio data or video data). Coder nine (the Research Coordinator) coded the interviews using all three formats simultaneously (see Figure 2). In contrast to a typical strategy of only coding selected text, video, or audio due to feasibility restraints (Bezemer & Mavers, 2011), each interview was coded in its entirety by a minimum of two coders for each format. This strategy helped conceptualize an understanding of this multimodal approach.

Coding is the most fundamental process in grounded theory (Strauss & Corbin, 1998). The process began in this study (Figure 3) by coders reading, listening, or watching all of the interviews to understand the participants’ experiences. Coders worked independently to complete open coding and create low-level categories from the interview data with coding decisions tracked in memos (Corbin & Strauss, 2015). Data were analyzed using a constructivist approach, in which open coding was applied in two sequential steps (initial and focused). Initial coding (often called line-by-line coding) consists of creating as many codes as needed, identifying the actions within them, and continuously comparing within and across sections of data (Charmaz, 2014). This reduces the likelihood of researchers reflecting their ideas in the data, maintains the focus on the participants’ perceptions of their realities, highlights the sensitizing concepts, and ensures that researchers systematically articulate their codes (Ong, 2012). Subsequent-focused coding enabled implicit concepts to be more explicit, led to the generation of categories, and developed larger analytical concepts (Charmaz, 2014). This systematic approach to coding “reduces the noise” or, in other words, makes the emerging themes more apparent and concrete (Jopke & Gerrits, 2019, p. 605).

After the initial phase of coding was completed, the research team conducted four 3-hour analysis meetings. Each independent coder shared their preliminary results of codes and categories and similarities and differences of interpretations across the three data formats were discussed. The research team compared the open coding results using 2-minute interview data segments to manage the large quantity of data. The segments were predetermined before the meetings so coders would be prepared to discuss their results. Six segments of interview data from five participants were used to compare the initial codes.

The research team designed a code sheet to help structure the data analysis meetings. As illustrated in Table 1 with an abbreviated example of one research team member’s code sheet for a single interview—the code sheet indicated the 2-minute interview segments, and the codes that the independent coder assigned in their initial analysis of that segment.

Each coder completed their code sheets (one for each participant assigned to them) and brought them to the data analysis meetings ready to discuss. Intercoder reliability was calculated during focused coding, using Fless’ kappa scores of agreement for each code. The intercoder reliability ranged from 0.62 to 0.81. These relatively high scores may be due to consensus in wording, even if slightly different terms were used. There was an average of four codes per 2-minute interview segment.
Figure 2. Coding assignments.

Figure 3. Coding process.

Table 1. Code Sheet Example.

<table>
<thead>
<tr>
<th>Interview Segment (Minutes)</th>
<th>Code 1</th>
<th>Code 2</th>
<th>Code 3</th>
<th>Code 4</th>
<th>Code 5, etc. . .</th>
</tr>
</thead>
<tbody>
<tr>
<td>6–8 Messages about being LGBTQ+ (bisexuality perceived as non-legitimate)</td>
<td>Inaccurate mainstream media stereotypes</td>
<td>Negative visibility/representation</td>
<td>Online blogging</td>
<td>. . .</td>
<td></td>
</tr>
<tr>
<td>10–12 LGBTQ+ images are stereotypical and limited</td>
<td>Gay male imagery is effeminate, fashionable, flamboyant, and dramatic</td>
<td>LGBTQ+ people often perceived in extremes—hyper-sexual or non-sexual</td>
<td>. . .</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14–16 Mainstream media sometimes has realistic representation of LGBTQ+ people</td>
<td>Messages about being LGBTQ+ can be simplistic/one-dimensional</td>
<td>Bisexual people perceived as indecisive or confused</td>
<td>Online media can help LGBTQ+ people cope with negative offline messages</td>
<td>. . .</td>
<td></td>
</tr>
<tr>
<td>18–20, etc . . .</td>
<td>. . .</td>
<td>. . .</td>
<td>. . .</td>
<td>. . .</td>
<td>. . .</td>
</tr>
</tbody>
</table>
During the data analysis meetings, the independent code sheets were merged into a coding table to make the code comparisons of data collection formats more visible. Each code table was organized by 2-minute segment. The research team reviewed the video recording of the interview segment and discussed their findings. The sustained focus of the coders was on triangulation through the multimodal analysis. In this article, triangulation refers to using multiple data sources of the transcript, audio, and video to comprehensively understand participants’ emotions and experiences (Carter et al., 2014; Denzin, 1978; Patton, 1999). Several steps were taken to enhance methodological rigor. Thick description (the extensive use of descriptive accounts and quotes), an audit trail (detailed recordings of the research steps and process), and member checking were utilized (Lincoln & Guba, 1985). The extensive notes, memos, and feedback from the large team of interviewers and coders were referenced throughout data analysis to confirm that codes and interpretations were grounded in the context of the participants’ experiences.

Coding Multiple Formats: Similarities and Differences

Notable similarities and differences were found in the coding of the same interview segments based on the data format used for analysis. Codes related to finding identities and community online, as well as offline and online safety issues, were not only similar—they were strengthened with the multimodal approach. For example, video data showed participants expressing strong non-verbal emotions (e.g., tears, enthusiastic body language) when discussing the positive, negative, and community-based aspects of their media engagement.

However, there were also key coding differences attributable to the data format during analysis. First, coders disagreed on the importance of participant affect in attributing meaning to statements. For example, the pragmatic Jeffersonian transcript from an audio file may include a bracketed note that the person is crying, but on the video they appear not to be tearful. Second, video coders noted discrepancies between verbal statements and body language missing from the transcript and audio coders’ analyses. Third, coders disagreed on the level of comfort and engagement of participants, particularly regarding distress when discussing traumatic experiences (e.g., violence).

Table 2 further illustrates the differences in modality related to the emotion generated as one participant discussed the impact of media messages on their mental health. While these issues may be mitigated through more rigorous transcription practices (e.g., multiple transcriptionists per interview), employing multimodal analysis helps to realize discrepancy and glean its meaning. The multimodal approach also revealed that fluctuations in tone and affect were most frequent when the semi-structured interviews expanded beyond the pre-developed questions to further probes and new lines of inquiry. The greater emotion present in these interview segments may be attributed to how semi-structured research instruments produce new knowledge beyond the initial conceptualization of the phenomenon under study.

Of the three data formats, video provided the most comprehensive data for analysis. It appeared to facilitate coder attunement to participants’ emotions in a more fulsome manner than transcript or audio, yet simultaneously inhibited attention to narrative details better plumbed by coding written transcripts. The video coders also found themselves relating more to participants on a personal level (either as reflections of self or of close relations) than the transcript and audio coders, who identified feeling more removed from participants. In this way, researcher positionality can vary across modes of coding and differing levels of reflexivity may be needed depending on the format in which analysts engage with data.

Implications

This paper highlights an example of multimodal coding applied to a constructivist grounded theory study with SGMY. The multimodal coding approach—using a combination of text, audio, and video—may be applicable across qualitative approaches to expand the analytical lens for qualitative research; generating a more accurate and nuanced conceptualization of the phenomena under investigation. Such an approach continues to incorporate transcripts, which help with narrative details and bolster the specifics of the analysis as compared to audio and visual data formats, while also capturing key emotions and non-verbal cues that can be missed in verbatim transcripts.

Researchers considering multimodal analysis may weigh the costs (e.g., more coders, greater time invested) and benefits (e.g., potentially greater depth) alongside the value of this approach to their worldview and format of data collection. This approach is potentially more useful for teams where numerous coders can be assigned distinct data formats to code and then discuss. The triangulation of these formats can help account for the presence of emotions (e.g., excitement, distress) and consider how these emotions as data can inform the analysis of the overall participant experience and contribute to developing a holistic interpretation of data. As illustrated in the example presented herein, this approach is particularly well-suited to semi-structured interviews due to the possibility of probing and exploring unanticipated topics that may benefit from a multimodal frame to develop a holistic understanding. More open interview formats may benefit similarly from this approach. In contrast, it may be less useful in more closed interview guides, such as a content analysis of a policy or program’s impact.

Drawing on a constructivist paradigm, scholars may be more likely to utilize multimodal analyses to enhance their perspective that the resulting research representations are constructed. However, researchers from other paradigms may also find this approach useful. For instance, researchers using a post-positivist paradigm may find that a multimodal coding contributes to greater rigor and trustworthiness of data.
Table 2. Comparison of Codes Relating to Emotion by Data Type.

<table>
<thead>
<tr>
<th>Minute</th>
<th>Sample Quote: Media Messages and Their Impact</th>
<th>Engaged?</th>
<th>Positivity</th>
<th>Negativity</th>
<th>Distress/Sadness</th>
<th>Reflective</th>
<th>Other Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>22</td>
<td>Basically now and beforehand, the biggest stories that I heard that were really sad and depressing are like—obviously the Brandon Teena story, that’s really scary. And then there’s Gwen, Araujo . . . That was really sad as well. And it’s kind of like, oh my god that kind of thing could happen to me. And it’s just sort of like—I just don’t want to present myself in a certain way that could get myself killed or whatnot. So actually seeing that, I was actually—I wasn’t sheltered, but I just wouldn’t go anywhere for like 2 years. And during high school, I just didn’t go to school. And I basically home-schooled myself because I just didn’t want to go out and have to deal with people and whatnot. And also—Even you hear in the media about 12-year-olds that are committing suicide because they’re gay. It’s just kind of like, that’s really young of an age to committing suicide. They haven’t even lived their life or anything yet or even tried living their lives. And it’s hard, that they feel like they have to die, basically, in order to just end their pain at such a young age. I was just—I was so fearful. I just thought that I would die if I walked outside my house. Somebody would see me and then they’d know me or whatever, and just kill me. So that was really stressful. But I coped by being basically online. Like I spent so many hours online in high school just looking up different resources and about trans people, and where to go for help, and they had support groups and whatnot. So that kind of gave me hope. And also I just kept a journal of all my feelings.</td>
<td>Transcript</td>
<td>Yes—Responds to each question (at times with lengthy explanations)</td>
<td>Yes—Verbally expressed positivity just when discussing online support</td>
<td>Somewhat—Words like stressful and difficult were used</td>
<td>Somewhat—Used the word sadness</td>
<td>Somewhat—specifically when discussing how what they heard about in the media could happen to them</td>
</tr>
<tr>
<td>23</td>
<td>Audio Yes—Answers quickly with an excited tone</td>
<td>No—Voice did not reflect positivity at anytime</td>
<td>Yes—Voice sounded flat and at times irritated</td>
<td>Yes—Eyes and facial expression looked concerned</td>
<td>Yes—Some pauses when speaking about their experiences and voice cracked occasionally</td>
<td>Yes—Took pauses when responding</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>Video Yes—Leans forward and makes eye contact with the interviewer</td>
<td>Somewhat—Smiled and became animated when discussing online coping and hope</td>
<td>Yes—Eyes and facial expression looked concerned</td>
<td>Yes—but then mild—Became almost tearful and looked upset initially then face became more fixed as they continued to speak</td>
<td>Yes—Sat back and sometimes pursed lips when reflecting on questions</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
interpretation than a unimodal analytical approach. As in this study, the use of technologies can capture the nuance of that construction and enable collaboration between researchers and participants (Pink, 2013). Aligning with the constructivist grounded theory, the use of multimodal analytic strategies allowed for constant comparison and triangulation between the three types of data, the relationships between interview concepts, and the diverse social locations of the coders (Vogl et al., 2019). Researchers from a transformative or participatory paradigm may also see value in this multimodal approach. It can accommodate a range of coders who may have different levels of reading, aural, or visual comprehension.

Audio recordings provided further insight into participant inflection (e.g., tone, pace) without the additional stimulus of video. The video was the most comprehensive data format for analysis. Video coders appeared to capture youths’ emotional ranges and communication strategies better than the transcript or audio coders. This may be a particularly important consideration for marginalized populations who may not be as comfortable communicating verbally. Further, research could explore the ways in which the participants use the technologies to highlight their perspectives and researchers use them to further the collaborative relationship necessary for effective data collection (McInroy, 2016; Pink, 2013).

Focusing on the experiences of SGMY in a multimodal framework illuminates the complexity that can remain unexplored when limited to individual perspectives and singular sources of data collection. Multimodal approaches help capture the similarities and differences of multiple participants and uncover the meanings they ascribe to those experiences and the ways they communicate about them (Kendall et al., 2009). Multimodal coding can potentially facilitate more profound insights into the processes that would not be clear using one approach in a non-triangulated analysis. For coders who identified as peer researchers, disagreement in interpretation aligns with research showing that peer identities do not necessarily align with shared lived experiences (Eaton et al., 2018; Marshall et al., 2012). More than mere description, a multimodal process should encourage researchers to dive deeper methodologically to explore both contradiction and agreement in an approach to self-reflection, and emerge with a constructed perspective on a phenomenon (Vogl et al., 2019).

Limitations

Although participants were aware of the multimodal approach to data collection as part of the informed consent, they may not have been aware of the multimodal approach to analysis, although randomized, and may have preferred to express themselves in different ways. Further, although a large number of coders were part of the analysis process, their interpretations may not have fully captured participants’ intentions. Future studies may consider including additional coders who review all three data sources. Despite such challenges, this study advances a more nuanced understanding of the similarities and differences concerning multimodal data collection and analysis for a marginalized population. As this approach resulted in a more holistic approach to data analysis that may have captured additional perceptions of and by the participants, it is hoped that these approaches will encourage future research initiatives that further explore the integration and approach to such investigations.

Conclusion

This paper presents a unique multimodal coding approach that incorporated triangulating three data formats (transcript, audio, and video) in its analysis of a qualitative study with SGMY. This approach enriched our study by illuminating emotion and affect in audiovisual formats and permitting comparison with codes from the textual transcript. Leveraging technology advancements for data analysis in multimodal approaches may promote richer analyses and offer an opportunity for comparing of coder interpretations across formats. Constructivist grounded theory, alongside CAQDAS software, provide a framework for independent coding and data synthesis that can triangulate formats to better understand the phenomena under investigation.

Acknowledgments

Thank you to Maria Staszkiewicz for assistance with this article, and thank you to our youth participants.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was funded by Insight Development and Partnership Grants from Canada’s Social Sciences and Humanities Research Council (SSHRC). SLC is a Canada Research Chair. ADE holds a salary award from the Ontario HIV Treatment Network (OHTN).

ORCID iD

Shelley L. Craig https://orcid.org/0000-0002-7991-7764

References


and Education, 19(2), 166–178. https://doi.org/10.1016/j.linged.2008.05.005


