


“Doors Started to Appear:” A Methodological Framework for Analyzing Visuo-Verbal Data Drawing on Roland Barthes’s Classification of Text-Image Relations

International Journal of Qualitative Methods
Volume 21: 1–12
© The Author(s) 2022
DOI: 10.1177/16094069221084433
journals.sagepub.com/home/ijq


Huw Grange¹  and Olaug S. Lian¹

Abstract

Following a “visual turn” in qualitative methods, photographs and other forms of visual expression are increasingly used in conjunction with verbal data in social science research. According equal status to visual and verbal artifacts, however, poses significant methodological challenges. “Photo elicitation” methods, which typically privilege participants’ interpretations of photographs over the photographs themselves, have dominated. This article answers calls for greater reflection on and transparency in the analysis of data across multiple modes of expression. Building on previous approaches, we propose an analytical framework for interpreting visuo-verbal research data that draws on Roland Barthes’s tripartite classification of text-image relations into “illustration,” “anchorage,” and “relay.” We explore how our framework can be put into practice by applying it to photographs and written texts generated as part of the “Living with an eating disorder during the COVID-19 pandemic” project, focusing on three settings represented by participants: the hospital ward, the home, and natural environments. We subsequently reflect on some of the strengths and limitations of our framework in light of its application and with respect to established approaches to analyzing visuo-verbal data. Our framework of Text-Image Relations Analysis enables researchers to explore text-image relations as constitutive of meaning without privileging one semiotic mode over the other. As with all qualitative researcher, however, careful delineation of the meaning-making roles of participants and researchers is key.

Keywords

photo narrative, narrative analysis, photo elicitation, photovoice, narrative research, methods in qualitative inquiry

Introduction

The “visual turn” in qualitative methods has seen increasing use and acceptance of visual artifacts, most notably photographs, in social science research (Bell, 2010; Pain, 2012). Two photographic methods are routinely distinguished (Glaw et al., 2017). In autophotography, firstly, participants take photographs (or at least stage them) and researchers analyze these artifacts *as data* (Glaw et al., 2017; Thomas, 2009). In photo elicitation, meanwhile, photographs (which may or may not have been created by participants themselves) are used to develop, inform and enhance *subsequent data collection*, for example, through interviews (Bell, 2010; Harper, 2002; Pain, 2012; Platzer et al., 2021). In practice, elements of

autophotography and photo elicitation are often combined. This is the case, for example, with the well-established participatory action research methodology Photovoice, in which the meaning of participant-generated photos is explored by the participants themselves in group dialogue sessions, at least in its original conception (Wang & Burris, 1997).

¹Department of Community Medicine, UiT The Arctic University of Norway, Tromsø, Norway

Corresponding Author:

Huw Grange, Department of Community Medicine, UiT The Arctic University of Norway, Postboks 6050, Langnes, Tromsø 9037, Norway.
Email: huw.r.grange@uit.no



Creative Commons CC BY: This article is distributed under the terms of the Creative Commons Attribution 4.0 License (<https://creativecommons.org/licenses/by/4.0/>) which permits any use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

It was the perceived “absolute and unqualified objectivity” (Strand, 1917, p. 524) of the camera that appealed to early adopters of visual methods (Pink, 2021). Since the 1960s, however, researchers working within relativist and constructivist paradigms have recognized the camera’s “twin capacities, to subjectivize reality and to objectify it” (Sontag, 2002, p. 178). In this case, a photograph is not considered an objective record of reality, but rather an artifact whose meaning is co-produced by photographer and viewer. The photographer selects their subject, composition, frame, lighting, time exposure, etc. (Berger, 2013; Chaplin, 1994), and in doing so lends their “point of view” (Thompson, 2003, p. 7). In the act of interpretation, moreover, each viewer brings their own “point of view” to an image insofar as the visual itself (i.e., *what* the human eye is physiologically seeing) is not the same as *how* we see what we see (Rose, 2016). Seeing is an experience, and people, not their eyes, see: “there is more to seeing than meets the eyeball” (Hanson, 1958, p. 7). Today, in an age of digital reproduction, photographs are also readily modified and transported to new contexts in which they are seen and interpreted in a new light. Photographs are “semantically promiscuous” (McQuire, 1998, p. 173), fragments whose “moorings come unstuck” over time, leaving them “open to any kind of reading” as they drift away (Sontag, 2002, p. 71).

Mindful of the challenges of interpreting visual artifacts, researchers working within a relativist or constructivist paradigm have privileged photo elicitation over autophotographic approaches, focusing their analysis on participants’ own interpretations of images (Brown & Collins, 2021; Chapman et al., 2017; Gleeson, 2011; Murray & Nash, 2017). Even with increasing use of visual methods, then, the visual has often been subordinated to the verbal (Chaplin, 1994). This not only risks overlooking a valuable source of knowledge (Chapman et al., 2017), but also ignores similar difficulties with interpreting verbal texts, which can also be understood to be “inherently polysemic” (Knowles & Sweetman, 2004, p. 13). Those scholars who have treated participant-generated photographs *as data* have been criticized, meanwhile, for refraining from reflecting upon or reporting how they went from image to findings (Catalani & Minkler, 2010).

This article builds on attempts to develop an approach to analyzing multimodal data without privileging one semiotic mode over another (Brown & Collins, 2021; Chapman et al., 2017; Gleeson, 2011). Answering calls for greater reflection on and transparency in analysis of visuo-verbal data (e.g., Catalani & Minkler, 2010), we begin by presenting a methodological framework that incorporates systematic investigation of text-image relations into the analytic process by drawing on the foundational work of Roland Barthes (Barthes, 1977a, 1977b) in semiotics. We then explore how our methodologically innovative approach can be implemented with reference to eight visuo-verbal illness narratives generated as part of a study into experiences of living with an eating disorder during the COVID-19 pandemic. Finally, in our discussion we highlight some of the strengths and limitations

of our framework in light of its application and with respect to established approaches to analyzing visuo-verbal data.

Analyzing Visuo-Verbal Data

Attempts to analyze visual data as part of a multimodal dataset have typically begun with the observation that, for all their differences, analyzing visual material has much in common with analyzing verbal material (Ritchie et al., 2014). For Gleeson (2011), for example, interpretation of pictures is akin to that of words, that is, “basically the same process of bringing one set of texts to bear on another in order to make meaning” (p. 314); for Chapman et al. (2017), the processes of coding images and texts are, essentially, “the same” (p. 814); and for Sellers (Braun & Clarke, 2021), a photograph can be treated in the same way as a transcript. Use of the same analytic method (e.g., content, thematic and narrative analysis) across semiotic modes has been advocated (Banks, 2018; Glaw et al., 2017; Rapport et al., 2007). Alternatively, “synergistic” methods may be found through experimentation (Drew & Guillemin, 2014, p. 63).

Some scholars have adopted what we might call a “holistic” approach to visuo-verbal data, whereby the visual and the verbal are merged into a single multimodal unit from the outset (Burles & Thomas, 2014; Lian & Rapport, 2016; Rose, 2016; Wilde et al., 2020). Others insist on analysis of visual and verbal data independently prior to analysis of the multimodal unit. Where codes are generated separately for images and words, these may be combined as part of a gradual shift from descriptive towards interpretive analysis (e.g., moving towards themes in thematic analysis). As Murray & Nash (2017) and Glaw et al. (2017) note, the rationale and procedure for combining codes are not always discussed thoroughly. For some, however, it is an additive process (a combined code needs only to have been ascribed to data in one mode, e.g., Brown & Collins, 2021); for others, it is a corroborative process (a combined code must have been ascribed to data in multiple modes, e.g., Chapman et al., 2017).

Studying the relationship between visual and verbal data is advocated (Rose, 2016; Pink, 2021), but seldom practiced. Among the rare examples of a relational stage built into the analytic process is Oliffe et al.’s (2008) formal “layered” analysis (drawing on Dowdall & Golden, 1989). They examine congruity between photographs and written texts as part of their “Review” phase. Similarly, in her “multiple text analysis,” Keats (2009) calls for intratextual (within case) consideration of the connections, parallels and differences between data in different modes. To the best of our knowledge, however, consideration has not yet been given to the systematic study of semantic relations between visual and verbal data prior to their analysis in combination.

Text-Image Relations in the Work of Roland Barthes

With the aim of developing a more rigorous approach to the study of participant-generated multimodal data, we turn to the

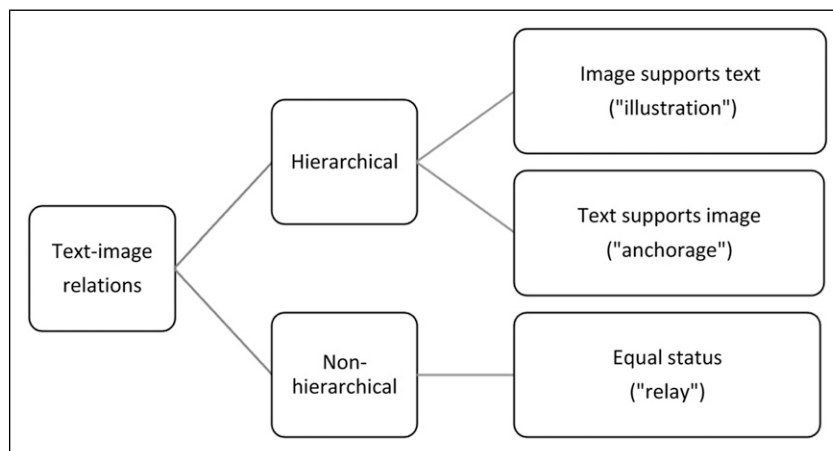


Figure 1. Barthes's Classification of Text-Image Relations.

foundational work in semiotics (the study of signs and sign systems) of Roland Barthes. Barthes (Barthes, 1977a, 1977b) remains the starting-point for scholarly discussion of text-image relations (Bateman, 2014). Although more complex classifications have since been developed by scholars of systemic functional linguistics to address logical gaps in Barthes's schema (Kong, 2006; Martinec & Salway, 2005; Van Leeuwen, 2005), the original provides the appropriate balance of rigor and flexibility for our present purposes.

Barthes's simple taxonomy of text-image relations recognizes three logical possibilities (see Figure 1):

What Barthes (Barthes, 1977a, p. 25) refers to as the most "traditional" relation is illustration. Here, an image supports a text, clarifying or "realizing" the written word, as is the case, for example, in Figure 3. Having dominated for centuries, illustration, Barthes (Barthes, 1977b) argues, has been overtaken by anchorage. In anchorage, a text supports an image by "anchoring" or elucidating its meaning, as in the case of photograph captions in newspapers (Barthes's key example). Barthes (Barthes, 1977b, p. 40) finds anchorage "repressive" insofar as it reduces the polysemy of images, directing the viewer to one particular interpretation among many. For an example taken from the data discussed later in this article, see Figure 4.

The final relation Barthes identifies is relay, in which the visual and verbal are accorded equal status. In cases of relay, words and images function as "fragments of a more general syntagm and the unity of the message is realized at a higher level" (Barthes, 1977b, p. 41). The whole, in other words, exceeds the sum of its parts. Barthes suggests (Barthes, 1977b) that cases of relay are relatively rare in static media, although Bateman notes (Bateman, 2014) that relay is recognized as being much more prevalent today. Comic strips are Barthes's prime examples. See Figure 5 for an example drawn from our data.

A Framework for Analyzing Visual-Verbal Data: Text-Image Relations Analysis

Drawing on the insights from Barthesian semiotics outlined above and on previous attempts to systematize data analysis

across multiple modes (notably Brown & Collins, 2021), we propose a framework for analyzing visuo-verbal data that incorporates systematic investigation of text-image relations. We stress that, as an analytical framework, our Text-Image Relations Analysis is flexible enough to be employed with a variety of analytical approaches (e.g., thematic analysis, grounded theory, IPA) and theoretical and philosophical perspectives. What we offer is simply a practical guide to acknowledging the relationship between data in visual and verbal modes and recognizing its potential for meaning-making.

According to our conception of Text-Image Relations Analysis, images and written texts are initially coded separately. The semantic relationship between image and text is then explored, for example, through the lens of Barthes's taxonomy (Barthes, 1977a, 1977b). Only then does integrated coding at the level of the multimodal unit take place, taking into account the text-image relations that have been identified. Special attention needs to be paid to cases of apparent incongruity between text and image (i.e., relay). Here, it may not be possible to interpret the multimodal whole in the absence of further data, and codes generated separately for the image and text should not therefore be combined. It is nonetheless important to note the form of text-image relation in case a pattern emerges.

As shown in Figure 2, analysis is an iterative process, in which images are compared and contrasted to images, texts to texts and visuo-verbal pairs to visuo-verbal pairs. The process of comparing and contrasting data elements takes place at two different levels: *within* cases and *between* cases. Codes should gradually be developed from descriptive to more analytical until conceptual themes have been generated that address the relevant research questions.

Applying the Text-Image Relations Analysis Framework

In this section we show how our framework of Text-Image Relations Analysis can be implemented, using data drawn

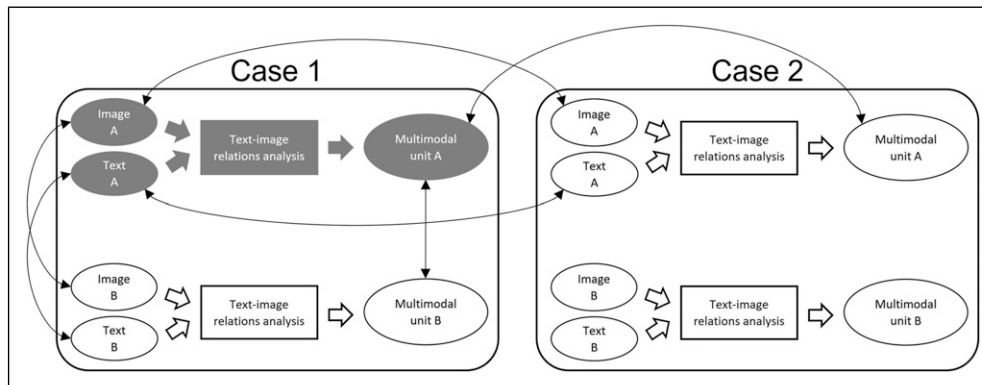


Figure 2. Text-Image Relations Analysis.

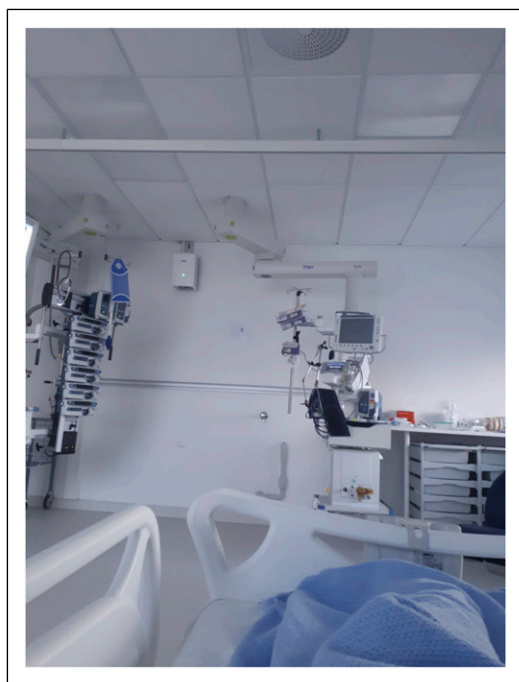


Figure 3. In the Somatic Hospital.

There have been a fair few admissions to both somatic and psychiatric hospitals during the “Corona year.”

from the *Living with an eating disorder during the COVID-19 pandemic* project. This project will be described more fully elsewhere. Here it is sufficient to say that participants from the UK and Norway submitted visuo-verbal illness narratives (each comprising up to 10 photographs with accompanying free-length caption-texts) about their experiences of living with an eating disorder during the first year of the COVID-19 pandemic (March 2020 to March 2021). Ethical approval for the project was obtained from the Norwegian Centre for Research Data (ID: 945186). Participants were informed of the aims, methods, and potential risks of taking part in the study and provided written informed consent in advance, including with regards to use of their photographic images in research



Figure 4. Hospital: The Turning Point.

This photo was taken in a German hospital, where I was kept after an unrelated accident due to my low blood pressure (caused by my Eating Disorder). A nurse commented on how I was “zu denn” (too thin) but all I could think about was how glad [I was] I did not have to eat because of the risk of illness from possible concussion, and how many calories I was “saving” by being hospitalized. It was this alarming thought process and event which led to my decision to move home.

publications. Quotations from written texts originally submitted in Norwegian have been translated into English by the lead author and checked by the co-author.

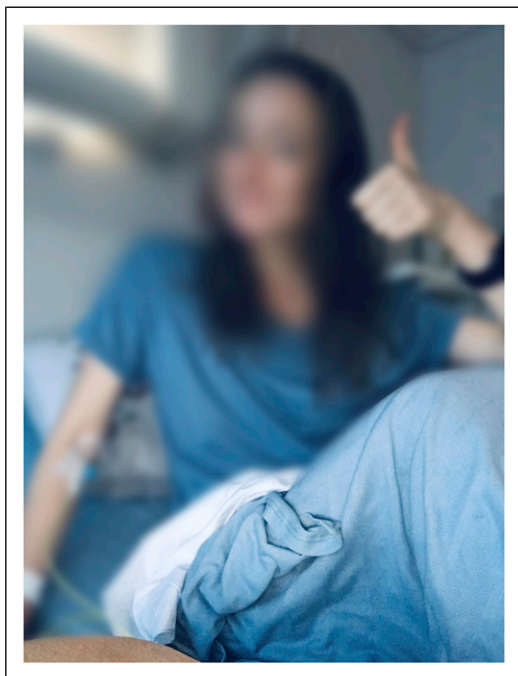


Figure 5. Contracted “Covid-19”.

I contracted Covid in April 2020 and was admitted to hospital with it. My history with anorexia impacted on how Covid affected me. My body was still in recovery from anorexia, my immune system was still building back up ... I unintentionally lost a lot of weight when I had Covid. This triggered the anorexia psychopathology ... a relapse was triggered.

During our analysis we observed that text-image relations varied according to the setting represented by participants. The nine visuo-verbal pairings we present here, then, are organized around three different settings: the hospital ward, the home, and natural (or semi-natural) environments.

Hospital Wards

Three photos (see Figures 3, 4, and 5) submitted by different participants represent experiences of inpatient treatment during the pandemic. Despite a commonality of subject—the participant’s hospital bed (Radley & Taylor, 2003)—each entertains a different relationship with the text that accompanies it. These relationships correspond to the three different text-image relations outlined in Barthes’s taxonomy:

The first photo-text pairing (see Figure 3) provides an example of illustration: the text explains that the photo captures a moment that has recurred, depicting one of “a fair few admissions.”

In Figure 4 the text embeds the photo within a broader narrative. This participant uses the caption-text to explain the reasons for her hospitalization (an “accident due to my low blood pressure”), her thoughts at the time (“all I could think about was ... how many calories I was ‘saving’”), and what happened next (“It was this ... which led to my decision to move home”). The dominant text-image relation here is anchorage: the text serves to situate the specific moment

captured by the photo in time and to pinpoint what is meaningful about this image for the participant (what makes it the eponymous “turning point”).

Compared to the other two figures, Figure 5 presents a more complex case. The image (partially pixelated, with the participant’s permission) depicts the participant sitting up in bed, smiling directly at the camera and making a “thumbs up” sign. Once again, the caption-text serves to anchor the image within a narrative, establishing that this photo was taken while the participant was hospitalized with COVID-19. She points to her eating disorder history as a factor that contributed to the severity of her COVID-19 infection and holds COVID-19 responsible for a subsequent anorexia relapse. There is an incongruity here, however, between image and text. The positivity connoted visually by the participant’s smile and thumbs up contrasts with the seriousness of the verbal narrative in which the image is embedded, one that tells of COVID-19 complications and eating disorder relapse. This is a “demand” image that requires something from the viewer (Kress & Van Leeuwen, 2021), in this case challenging us as readers and viewers to make sense of its relationship to the verbal text. In light of the caption, we returned to the image and asked ourselves whether it depicted the participant’s determination to preserve a positive outlook in the face of illness, or perhaps critiqued a culture of “positive thinking” (Ehrenreich, 2009), which, in the participant’s case, may have done little to help. Image and text combine here, in a relationship of relay, to form a multimodal unit that is suggestive of meaning beyond that connoted by its constituent parts. There may, however, be several contenders for what that “more general syntagm” (Barthes, 1977b, p. 41) might be.

An additive approach to coding visuo-verbal data may prove adequate for cases of illustration and anchorage. To the images in Figures 3 and 4, for example, we initially ascribed a code of “Hospital ward,” with regards to their accompanying texts we opted for “Hospitalization.” When moving to analysis of the multimodal unit, reconciliation of these separate codes to form combined codes did not pose significant problems. Analysis of the third photo-text pairing (Figure 5), however, required more careful consideration. We initially ascribed codes of “Hospital ward,” “Smiling,” and “Making thumbs up” to the image, and “Hospitalization” and “Worsening ED symptoms” to the text. But to simply add these together, or select those common to text and image, would have been to overlook supplementary meaning generated by the particular ways in which semiotic resources have been combined by the participant. We initially kept the codes that we had generated for the text and photo separate, but we also noted the particular text-image relation (“image more positive than text”) in case a pattern emerged as we compared and contrasted this visuo-verbal pairing to others in the dataset. In this case a pattern did emerge, and the relevant multimodal sets were given the *in vivo* code “Behind the smile I was really struggling,” a phrase that the participant who produced Figure 5 used elsewhere in her submitted work.



Figure 6. *Counting.*

I notice that I am counting the calories more now that I am at home a lot and both have access to scales and can also sort of keep track and be in control, by making all my own food in my own kitchen.

Illustrating Home

Participants living alone during the pandemic repeatedly linked spending more time at home to exacerbated eating disorder symptoms, including increased body checking (looking in the mirror, pinching parts of the body), calorie counting, bingeing, and self-induced vomiting. Their photos tend to illustrate activities described in the accompanying texts. Some capture activities as they were taking place, for example, mirror selfies depicting body checking. More commonly, photos depict objects that are “indexical” of those activities (Ledin & Machin, 2018), for example, depicting a bin overflowing with food wrappers to represent bingeing, or a toilet to represent purging.

Figure 6 provides an example of an image depicting objects from which we infer an activity. The photo is a straightforward depiction of kitchen scales, packets of oats and a sugar substitute, and a scrap of paper with calculations on it. A straightforward text accompanies it: the participant explains that more time at home has led to more calorie-counting. In terms of text-image relations, this is a clear case of illustration: the photograph captures a single instance of what the text describes as habitual practice.

That participants repeatedly had recourse to illustration (depicting one instance among many described in words) rather than anchorage (with its turning points) was significant. This pattern resonated with descriptions of homelife during the pandemic as monotonous, for example, “Being in lockdown meant that there was very little change from day to day.”



Figure 7. *Distraction Techniques or Lockdown Hobbies.*

While a lot of people took up new hobbies to fill their new-found free time in lockdown, I had an ulterior motive. I took up sewing as a distraction technique early in my recovery. I started with a cross stitch kit, making a floral bookmark and glasses case, and then went on to buy my own embroidery hoops and finally use the threads I inherited from my grandma when she died. I keep my sewing things next to me when I eat, and mostly pick it up straight after finishing my meal or snack to keep my hands and brain busy and the guilty thoughts at bay.

The text-image pair in Figure 7 is among the few representations of homelife submitted by participants that go beyond illustration. In her writing, the participant describes taking up sewing “as a distraction technique” to keep her “hands and brain busy and the guilty thoughts at bay.” At first glance, the photo may appear to be a straightforward illustration, a depiction of the sort of embroidery mentioned in the text. On closer inspection, however, details emerge. Both embroideries are food-related: the carton of fries on the right obviously so, the larger hoop on the left taking the form of a pictorial diary with several food-related items depicted (a hamburger, cake, sweets, a frying pan, etc.). The image in Figure 7, then, directs the viewer to thoughts about food and eating, even as the text reports that the very reason the participant embroiders is to distract herself from such thoughts.

The incongruity between text and image here points to a relationship of relay. There are several ways we might make sense of the multimodal unit here. Perhaps the participant’s attempts to distract herself from obsessive food-related thoughts with needlework were not always successful, and the photo allowed her to express this difficult subject more easily than in words. Perhaps, on the contrary, she found it perfectly easy to disassociate the act of embroidering from the subject of the embroidery itself. Perhaps food was less prominent as a theme in her other needlework. Why she chose to pair text and image in this way might form the basis of a (sensitively handled) discussion in a follow-up interview (Chapman et al., 2017). In the absence of further data, however, it is difficult to privilege one interpretation of the multimodal whole over another. When

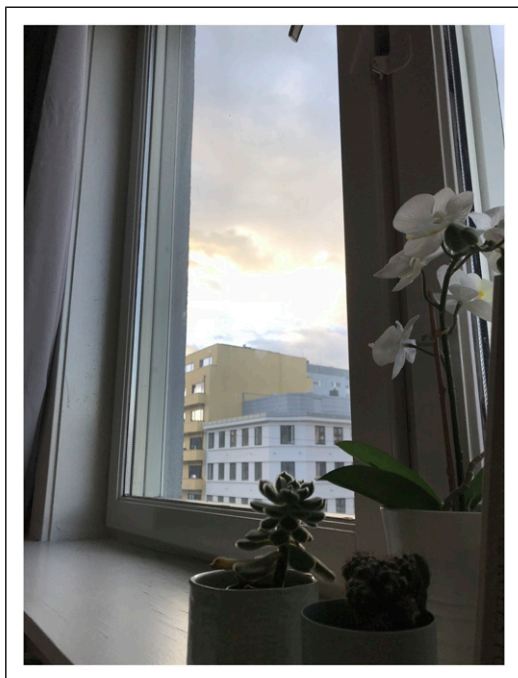


Figure 8. *Window Sill.*

I could finally stay in by myself, only eating food that was safe, counting calories, fasting as long as I wanted to and exercising as much as I wished. I could focus all my energy into my weight-loss project ... I sat at home in the corner by the window and planned shopping lists, meals, mealtimes, home workouts. I felt I had control, and with my strict regime it did not take long for me to lose a lot of weight.

performing integrated coding, then, we opted to keep the codes we generated for the image (“Depiction of food”) and for the text (“Needlework as distraction technique”) separate. But we also coded the text-image relation (“Relay”), in case similar incongruities recurred elsewhere in our data.

Figure 8 provides a further example of a representation of homelife where the relationship between text and image is one of relay. The photo depicts a window with succulents and a white orchid resting on the sill, with the viewer’s gaze directed through the windowpane towards apartment blocks on the opposite side of the street. We noted the similarity of this photograph to some of the hundreds of thousands of others circulating on social media at the time, when people under “lockdown” restrictions around the globe were encouraged to share the view from their window in a bid “to connect and escape” (e.g., the *View from my window* project, Duriau, 2021). We also considered it in the light of photographs of windows taken as part of a study into illness experiences of people with medically unexplained long-term fatigue, where the window functioned as a site of information exchange and connection (Lian & Lorem, 2017). In the text in Figure 8, however, the participant reveals that this photo is significant for her not because of the glimpse of the outside world it affords, or for that matter because of the houseplants, but because it was taken from the corner of the room where she habitually planned her weight-loss project. With lockdown providing the perfect

conditions for her eating disorder to flourish, this participant reports that, at the time, she had no interest in gazing through the window; in connection or escape.

This multimodal unit left us wondering whether the participant, in combining text and image in this way, intended to contrast her experience of the pandemic with that of others (as, for example, the participant in Figure 7 contrasts her reason for taking up a new hobby with that of others), or whether this was simply a window through which she looked, uninterested, as she planned her weight-loss regime. Again, in the absence of further data, we kept codes generated for the text and photo separate, but signaled the text-image relation at the multimodal level in case a pattern emerged (e.g., recurring use of photos to convey contrasting experience of others).

Getting Out

Compared to the inside spaces they depicted and described, participants for the most part represented outside spaces much more positively. In their writing they point to natural and semi-natural environments such as woodland, fields and parks as “therapeutic landscapes” (Gesler, 1992). They report being able to escape negative thinking patterns in these places, leaving them feeling “grounded,” “connected,” and more “mindful.” These were relaxing environments in which they felt able to talk to others about their illness and to challenge destructive eating habits. The beauty of nature, moreover, inspired hope and increased motivation for recovery. In the words of one participant: “Nature is a really special form of therapy.”

While this corpus did contain photographs that illustrated routine encounters with nature, and texts that anchored images of natural landscapes within a narrative, a majority of the images depicting outside spaces (12, split across four participants) were related to caption-texts through metaphor, wherein the characteristics or qualities of one object are applied to another (Lakoff & Johnson, 1980; Switzer, 2019). Barthes did not explicitly consider how metaphor might fit into his schema of text-image relations (Bateman, 2014). Insofar as a text that explicitly confers a metaphorical interpretation on an image (or part thereof) reduces its polysemic potential, it can be considered an example of anchorage. In many cases, however, a text will develop a metaphor from an image without explicitly pointing to the image as its source, or without elucidating that metaphor in full. Here we are dealing with a relation of relay: the reader/viewer is left to move back and forth between semiotic modes, looking for ways in which aspects represented in one can be transferred to the other.

In Figure 9 the participant does much of the interpretive work for us. The text indicates two stages of figuration. First, cycling has come to symbolize “freedom” for this participant. Second, now that gyms are closed, he has come to realize that his pre-pandemic exercise regime was “compulsive and rigid.” When cycling, by contrast, he does not feel “locked down” at



Figure 9. *Transported.*

One of the first things I did in lockdown was adapt my exercise routines. I was very anxious about the gym closing, as I used to go every day. I also recognized however that going to gym had become compulsive and rigid. Finding cycling again helped me remember that movement should be about freedom, not constraints. We had wonderful weather during the first lockdown, and heading out on my bike made me feel transported away from the mundane four walls of my room. I could both escape and connect at the same time and I realized I was not locked down, after all.

all, even while living under restrictive social distancing measures. In addition to the *FREEDOM IS CYCLING* metaphor (to articulate this metaphor in the conventional way, with “target” preceding “source”), we have a *COMPULSIVE EXERCISE IS LOCKDOWN* metaphor. With a figurative interpretation for the photo established by the text, we might return to the image and examine how visual elements that are not explicitly mentioned in the text contribute to the metaphor. The almost cloudless sky and vast, open fields stretching to the horizon, for example, could be said to supplement the participant’s symbolization of cycling as freedom.

Figure 10 presents a more complex case of metaphor: compared to the previous example, the reader/viewer is left to do more of the interpretive work themselves. The first photo in this participant’s series depicts a woman (either the participant or a friend) standing on a wooded hillside, her back turned as she looks towards a misty horizon and the valley below. The photo’s title (*One of several hikes with good friends*) makes reference to hiking and anticipates a text-image relation of illustration (we expect the image to depict one instance of many cases reported in the text). The caption-text itself, however, makes no mention of hiking. Instead, it points to gym closures during the pandemic as playing a significant role in the participant’s eating disorder recovery. Whereas, in Figure 9, pre-pandemic compulsive exercise in the gym is explicitly contrasted with a form of exercise that permits contact with nature, in Figure 10 seems to establish an implicit opposition through a text-image relation of relay.

The text here does not “anchor” a metaphorical interpretation to the image. Rather, the loose, unwritten bond between text and caption prompts the viewer to search for a figurative meaning in the photo based on their knowledge of images’

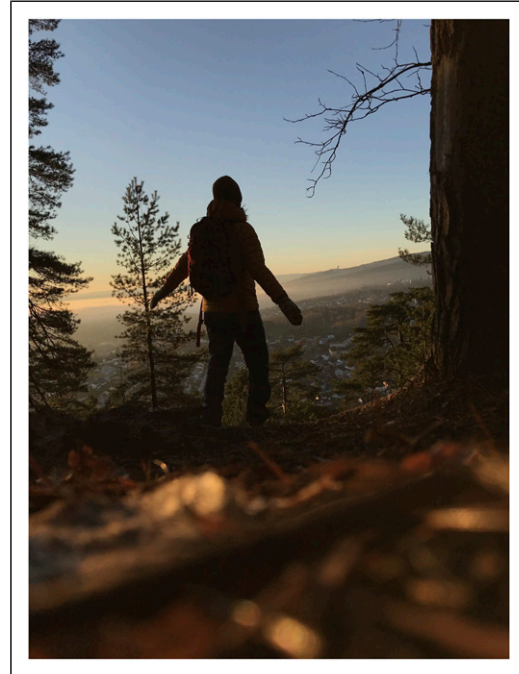


Figure 10. *One of Several Hikes with Good Friends.*

The start of the pandemic went pretty well for me. I went home and was there a month or so and in the following months I was sociable with friends and started a little project all by myself where I decided to recover completely from ED. I worked hard every day and managed to make absolutely amazing changes that I am really proud of. The fact that gyms closed helped me A LOT to finally take a break from the compulsive relationship I had with it.

“canons of use” (Ledin & Machin, 2018, p. 47). Not all viewers will see the same thing. Working in a Norwegian context, we, for example, found the protagonist gazing towards a mysterious horizon to be reminiscent of Romantic and Neo-Romantic landscape painting (Theodor Kittelsen’s *Soria Moria*, for example). Connoting hopefulness and a spirit of adventure, for us this photo visually conveys the new beginning—the participant’s decision to work towards a full recovery from her eating disorder at the start of the pandemic—described in the text.

In Figure 10, the caption-text may make no reference to getting outside, but we are at least supplied with a title that anchors the natural landscape depicted in the photo at a literal level. Figure 11, however, leaves us with many more unanswered questions. The photo is a close-up taken near ground level of a colorful little door that has been placed among nettles at the foot of a tree-trunk. The caption-text gives no indication of the circumstances in which this photo was taken: no mention is made of a walk in the woods, let alone whether the door formed part of an art installation (e.g., Dinky Doors, 2021) or was placed there by the participant herself. The text simply metaphorizes one element of the image (*POSITIVE EFFECT OF TREATMENT IS A DOOR APPEARING*). Prompted to transfer qualities of the (visual) source to the (written) target, we as



Figure 11. *Doors Appeared.*

I began to engage in quite intense ED community treatment. Accepting help was hard, accepting I needed the help was even harder. But, as soon as I did, doors started to appear.

readers/viewers returned to the image and considered how the door's diminutive size might be suggestive of incremental progress, how the nettles could indicate the challenges to and precariousness of recovery, and the participant might be using this unexpected and delightful door to communicate that she found the benefits of community treatment to be similarly unexpected and delightful.

Insofar as the three caption-texts presented in this section all derive a metaphor from their corresponding photos, the text-image relation at work is that of relay. The point here is not that all readers/viewers will develop a metaphor in the same way; others will interpret the text-image pairings differently, including, perhaps, the participants themselves. What matters is that a relation of relay sparks metaphorical thinking in the reader/viewer.

When coding at the multimodal level, then—conscious of the fragility and provisionality of our interpretations—we refrained from merging codes generated separately for texts and images. What we did do was ascribe an integrated code of “relay” to each pair. This allowed us to notice a pattern in participants’ usage of text-image relations: the natural and semi-natural spaces they photographed provided a source of metaphor through which they understood and expressed pathways to recovery. Participants pointed to nature as beneficial for their recovery. But by taking text-image relations into account in our analysis of the visuo-verbal illness narratives, we were able to uncover one of the reasons *why* it may have helped.

Discussion

According to the old adage, “a picture is worth a thousand words.” The reasons researchers like to work with participant-generated photographs or other visual artifacts alongside their words are well-rehearsed. Images have the potential to convey people’s experiences and life-worlds with great precision and emotional power and, moreover, where words prove to be inadequate (Berger et al., 1972). They capture “small, unintended details” that nonetheless reveal something significant about a participant’s lifeworld (Shannon et al., 2021, p. 118). In addition, they offer insights into the physical settings of people’s lives, serving to “re-anchor” study participants in physicality (Rugg, 1997, p. 2).

Triangulating visual and verbal data may also give the researcher greater confidence in their findings: “as soon as photographs are used with words, they produce together an effect of certainty ... Together the two then become very powerful” (Berger, 2013, p. 66). Words can authorize a single interpretation of an otherwise ambiguous photo, while a photo—“irrefutable as evidence” (Berger, 2013, p. 66)—lends authenticity to the words that accompany it, especially if we believe “the camera cannot lie” (Collier & Collier, 1967/1986, p. 8).

Combining more than one mode, however, can also create uncertainties for the researcher. A picture may be worth a thousand words, but those thousand words are likely to be different for different viewers. Even where the meaning-making potential of photographs is recognized and harnessed, then, the temptation is often to locate meaning ultimately in verbal interpretations rather than in images themselves (as in photo elicitation approaches).

Barthes’s (Barthes, 1977a, 1977b) classification of text-image relations helps us to understand why texts and images can be situated so differently by different researchers. Berger’s (Berger, 2013, p. 66) “effect of certainty” is fostered by texts and images tied by relations of illustration and anchorage. Analyzing visuo-verbal data, in these cases, may simply be a case of coding at the multimodal level immediately, or else combining codes developed separately for data in each mode. Where relay is the dominant relation, however—and there is reason to believe we should expect relay to be the dominant relation more often than not (Bateman, 2014)—that certainty is challenged. Relay trades on “the secret antipathy between the two modes of expression” (Hunter, 1987, p. 25). Rather than minimizing ambiguity, it opens it up. The reader/viewer is invited to make sense of the relation between image and text, and to interpret the supplementary meaning generated at the level of the multimodal unit.

Thus far, there has been little reflection in scholarship on what to do when we as researchers receive that invitation. When confronted with incongruities between visual and verbal data, Chapman et al. (2017) recommend pursuing further dialogue with participants. Likewise, Rapport et al. (2007) report that they are minded to take a photo elicitation

approach in a follow-up study to make sense of one dataset “undercutting” the other. In cases of relay, the temptation is to reassert the primacy of verbal data over the visual. Others hold fast to the principle of according equal status to visual and verbal modes: for Oliffe et al. (2008) it is the researcher’s responsibility to use context “to explain, rather than expose, what appeared to be incongruous details” (p. 534), while Brown and Collins (2021) advise researchers to adopt an open, critical-reflective stance to make sense of “differences, discrepancies, and contradictions” between forms of communication (p. 1287). As we noted in our discussion of the participant’s attempts to distract herself from food through crafts in Figure 7, however, it may be inappropriate (and at the very least feel profoundly uncomfortable) to assert an interpretation that conflicts with a participant’s own words, privileging the voice of the researcher over that of the participant. It may, in other words, be difficult to distinguish between “explanation” and “exposure.”

The key strength of the analytical framework presented here is that it constitutes a “middle way” between recourse to photo elicitation and imposition of researcher interpretation. Text-Image Relations Analysis provides researchers with a tool for analyzing data across modes of expression without privileging the verbal over the visual and without privileging the researcher’s voice over the participant’s. Meaning-making takes place *between* modes of equal status and *between* participants and researchers, who are accorded distinct, mutually enhancing roles. There are nonetheless several methodological challenges here. The semantic relation between a text and image may not be immediately apparent. Multiple relations may appear to operate simultaneously. Barthes’s tripartite classification that we have employed here, moreover, inevitably fails to capture the full range of possible text-image relations (Martinec & Salway, 2005). We need to be mindful, in other words, that identifying text-image relations is in itself an act of interpretation.

Confronted with cases of relay, the option a researcher will choose will depend on their ontological and epistemological assumptions about different modes of expression and the ways in which they combine to produce meaning. When working with multiple semiotic modes, it is important to resist the “effect of certainty” data triangulation can bring. We need, rather, to reflect deeply and sustainedly on how we are situating individual modes and their combinations in relation to truth and knowledge. This includes examining the limitations of the speech and written texts that continue to dominate as sources of data in research today (Reavey & Johnson, 2017). Combining visual with verbal data holds considerable, underutilized potential for gaining greater understanding of people’s life-worlds than words or images alone can deliver. To harness that potential, however, requires that we accord equal status to words and images, while also acknowledging and carefully delineating the meaning-making roles of participants and researchers in relation both to each semiotic mode and to their combinations (Drew & Guillemin, 2014).

Conclusion

Participant-generated photographs and other visual forms of expression are increasingly used by qualitative researchers alongside verbal research data. Images can complement words, for example by re-anchoring participants’ words in the physical world or by capturing “unintended details.” In addition, the configuration of words and images into a multimodal unit can yield insights into people’s experiences and life-worlds that transcend those afforded by data in individual semiotic modes alone. Analysis across modes poses significant methodological challenges, not least because multimodal data can introduce another layer of ambiguity into a dataset. Photo elicitation methods have therefore tended to predominate, and the potential of visuo-verbal data remains underutilized.

As a step towards harnessing that potential, we propose a framework of Text-Image Relations Analysis for interpreting visuo-verbal research data, drawing on Roland Barthes’s tripartite classification of text-image relations into “illustration,” “anchorage,” and “relay.” Application of our framework to eight multimodal illness narratives generated as part of the *Living with an eating disorder during the COVID-19 pandemic* project highlights a key advantage compared to existing approaches: introducing an additional element into the analytic process, it allows researchers to explore text-image relations as constitutive of meaning without privileging one semiotic mode over the other. As in all qualitative research, carefully delineating the meaning-making roles of participants and researchers is crucial.

Acknowledgments

With thanks to Rådgeving om spiseforstyrrelser (ROS) and SPISFO for their advice during the design and development of the *Living with an eating disorder during the COVID-19 pandemic* project, and especially to all the contributors.

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) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Huw Grange  <https://orcid.org/0000-0002-3791-1321>

References

- Banks, M. (2018). *Using visual data in qualitative research* (2nd ed.). Sage. <https://doi.org/10.4135/9781526445933>.
- Barthes, R. (1977a). The photographic message (S. Heath, Trans.). In R. Barthes, *Image, music, text* (S. Heath, Trans.) (pp. 15–31).
- Hill and Wang. Original work published 1961.

- Barthes, R. (1977b). Rhetoric of the image (S. Heath, Trans.). In R. Barthes, Image, music, text (S. Heath, Trans.) (pp. 32–51). Hill and Wang. Original work published 1964.
- Bateman, J. A. (2014). *Text and image: A critical introduction to the visual/verbal divide*. Routledge.
- Bell, S. E. (2010). Visual methods for collecting and analysing data. In I. Bourgeault, R. Dingwall, & R. De Vries (Eds.), *The SAGE handbook of qualitative methods in health research* (pp. 513–535). Sage. <https://doi.org/10.4135/9781446268247>
- Berger, J. (2013). Appearances. In J. Berger (with G. Dyer), *Understanding a photograph* (pp. 61–98). Penguin Books. Original work published 1982.
- Berger, J., Blomberg, S., Fox, C., Dibb, M., & Hollis, R. (1972). *Ways of seeing*. Penguin Books.
- Braun, V., & Clarke, V. (2021). *Thematic analysis: A practical guide*. Sage.
- Brown, N., & Collins, J. (2021). Systematic visuo-textual analysis: A framework for analysing visual and textual data. *The Qualitative Report*, 26(4), 1275–1290. <https://doi.org/10.46743/2160-3715/2021.4838>
- Burles, M., & Thomas, R. (2014). “I just don’t think there’s any other image that tells the story like [this] picture does”: Researcher and participant reflections on the use of participant-employed photography in social research. *International Journal of Qualitative Methods*, 13(1), 185–205. <https://doi.org/10.1177/160940691401300107>
- Catalani, C., & Minkler, M. (2010). Photovoice: A review of the literature in health and public health. *Health Education & Behavior*, 37(3), 424–451. <https://doi.org/10.1177/1090198109342084>
- Chaplin, E. (1994). *Sociology and visual representation*. Routledge.
- Chapman, M. V., Wu, S., & Zhu, M. (2017). What is a picture worth? A primer for coding and interpreting photographic data. *Qualitative Social Work*, 16(6), 810–824. <https://doi.org/10.1177/1473325016650513>
- Collier, J., & Collier, M. (1986). *Visual anthropology: Photography as a research method* (revised ed. with E. T. Hall). University of New Mexico Press. Original work published 1967.
- Dinky doors (2021). <https://www.dinkydoors.co.uk/>
- Dowdall, G. W., & Golden, J. (1989). Photographs as data: An analysis of images from a mental hospital. *Qualitative Sociology*, 12(2), 182–213. <https://doi.org/10.1007/BF00988997>
- Drew, S., & Guillemin, M. (2014). From photographs to findings: Visual meaning-making and interpretive engagement in the analysis of participant-generated images. *Visual Studies*, 29(1), 54–67. <https://doi.org/10.1080/1472586X.2014.862994>
- Duriau, B. (2021). *View from my window*. <https://viewfrommywindow.world/>.
- Ehrenreich, B. (2009). *Bright-sided: How the relentless promotion of positive thinking has undermined America*. Metropolitan Books.
- Gesler, W. M. (1992). Therapeutic landscapes: Medical issues in light of the new cultural geography. *Social Science and Medicine*, 34(7), 735–746. [https://doi.org/10.1016/0277-9536\(92\)90360-3](https://doi.org/10.1016/0277-9536(92)90360-3)
- Glaw, X., Inder, K., Kable, A., & Hazelton, M. (2017). Visual methodologies in qualitative research: Autophotography and photo elicitation applied to mental health research. *International Journal of Qualitative Methods*, 6(1), 160940691774821. <https://doi.org/10.1177/1609406917748215>
- Gleeson, K. (2011). Polytextual thematic analysis for visual data pinning down the analytic. In P. Reavey (Ed.), *Visual methods in psychology: Using and interpreting images in qualitative research* (pp. 346–361). Routledge.
- Hanson, N. R. (1958). *Patterns of discovery: An inquiry into the conceptual foundations of science*. Cambridge University Press.
- Harper, D. (2002). Talking about pictures: A case for photo elicitation. *Visual Studies*, 17(1), 13–26. <https://doi.org/10.1080/14725860220137345>
- Hunter, J. (1987). *Image and word: The interaction of twentieth-century photographs and texts*. Harvard University Press.
- Keats, P. A. (2009). Multiple text analysis in narrative research: Visual, written, and spoken stories of experience. *Qualitative Research*, 9(2), 181–195. <https://doi.org/10.1177/1468794108099320>
- Knowles, C., & Sweetman, P. (2004). Introduction. In C. Knowles, & P. Sweetman (Eds.), *Picturing the social landscape: Visual methods and the sociological imagination* (pp. 1–17). Routledge.
- Kong, K. C. C. (2006). A taxonomy of the discourse relations between words and visuals. *Information Design Journal*, 14(3), 207–230. <https://doi.org/10.1075/idj.14.3.04kon>
- Kress, G., & Van Leeuwen, T. (2021). *Reading images: The grammar of visual design* (3rd ed.). Routledge. Original work published 1990.
- Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. University of Chicago Press.
- Ledin, P., & Machin, D. (2018). *Doing visual analysis: From theory to practice*. Sage.
- Lian, O. S., & Lorem, G. F. (2017). “I do not really belong out there anymore”: Sense of being and belonging among people with medically unexplained long-term fatigue. *Qualitative Health Research*, 27(4), 474–486. <https://doi.org/10.1177/1049732316629103>
- Lian, O. S., & Rapport, F. (2016). Life according to ME: Caught in the ebb-tide. *Health*, 20(6), 578–598. <https://doi.org/10.1177/1363459315622041>
- Martinez, R., & Salway, A. (2005). A system for image-text relations in new (and old) media. *Visual Communication*, 4(3), 337–371. <https://doi.org/10.1177/1470357205055928>
- McQuire, S. (1998). *Visions of modernity: Representation, memory, time and space in the age of the camera*. Sage.
- Murray, L., & Nash, M. (2017). The challenges of participant photography: A critical reflection on methodology and ethics in two cultural contexts. *Qualitative Health Research*, 27(6), 923–937. <https://doi.org/10.1177/1049732316668819>
- Oliffe, J. L., Bottorff, J. L., Kelly, M., & Halpin, M. (2008). Analyzing participant produced photographs from an ethnographic study of fatherhood and smoking. *Research in Nursing & Health*, 31(5), 529–539. <https://doi.org/10.1002/nur.20269>
- Pain, H. (2012). A literature review to evaluate the choice and use of visual methods. *International Journal of Qualitative Methods*, 11(4), 303–319. <https://doi.org/10.1177/160940691201100401>
- Pink, S. (2021). *Doing visual ethnography* (4th ed.). Sage. Original work published 2001.

- Platzer, F., Steverink, N., Haan, M., de Greef, M., & Goedendorp, M. (2021). The bigger picture: Research strategy for a photo-elicitation study investigating positive health perceptions of older adults with low socioeconomic status. *International Journal of Qualitative Methods*, 20, 1-11. <https://doi.org/10.1177/160940692111040950>.
- Radley, A., & Taylor, D. (2003). Images of recovery: A photo-elicitation study on the hospital ward. *Qualitative Health Research*, 13(1), 77-99. <https://doi.org/10.1177/1049732302239412>
- Rapport, F., Doel, M. A., & Elwyn, G. (2007). Snapshots and snippets: General practitioners' reflections on professional space. *Health and Place*, 13(2), 532-544. <https://doi.org/10.1016/j.healthplace.2006.07.005>
- Reavey, P., & Johnson, K. (2017). Visual approaches: Using and interpreting images. In C. Willig, & W. Stainton Rogers (Eds.), *The SAGE handbook of qualitative research in psychology* (pp. 354-373). Sage. <https://doi.org/10.4135/9781526405555>
- Ritchie, J., Lewis, J., McNaughton Nicholls, C., & Ormston, R. (2014). *Qualitative research practice: A guide for social science students and researchers* (2nd ed.). Sage. Original work published 2003.
- Rose, G. (2016). *Visual methodologies: An introduction to the interpretation of visual materials* (4th ed.). Sage. Original work published 2001.
- Rugg, L. H. (1997). *Picturing ourselves: Photography and autobiography*. University of Chicago Press.
- Shannon, J., Borron, A., Kurtz, H., & Weaver, A. (2021). Re-envisioning emergency food systems using photovoice and concept mapping. *Journal of Mixed Methods Research*, 15(1), 114-137. <https://doi.org/10.1177/1558689820933778>
- Sontag, S. (2002). *On photography*. Penguin Books. Original work published 1977.
- Strand, P. (1917). Photography. *Seven Arts*, 2, 524-525.
- Switzer, S. (2019). Working with photo installation and metaphor: Re-visioning photovoice research. *International Journal of Qualitative Methods*, 18, 160940691987239. <https://doi.org/10.1177/1609406919872395>.
- Thomas, M. E. (2009). Auto-photography. In R. Kitchen, & N. Thrift (Eds.), *International encyclopedia of human geography* (pp. 224-251). Elsevier.
- Thompson, J. L. (2003). *Truth and photography: Notes on looking and photographing*. Ivan R. Dee.
- Van Leeuwen, T. (2005). *Introducing social semiotics*. Routledge.
- Wang, C., & Burris, M. A. (1997). Photovoice: Concept, methodology, and use for participatory needs assessment. *Health Education & Behavior*, 24(3), 369-387. <https://doi.org/10.1177/109019819702400309>
- Wilde, L., Quincey, K., & Williamson, I. (2020). "The real me shining through M.E.": Visualizing masculinity and identity threat in men with myalgic encephalomyelitis/chronic fatigue syndrome using photovoice and IPA. *Psychology of Men & Masculinities*, 21(2), 309-320. <https://doi.org/10.1037/men0000220>