Visual Culture Isn't Just Visual:
Multiliteracy, Multimodality and Meaning

Paul Duncum
University of Illinois at Urbana-Champaign

The central claim of this article is that contemporary cultural forms such as television and the Internet involve more than the perceptual system of sight and more than visual images as a communicative mode. Meaning is made through an interaction of music, the spoken voice, sound effects, language, and pictures. This means that even the recent shift to visual culture among art educators is insufficient to position art education within a reconceptualized, broad definition of communications. To be relevant to contemporary social practice, art education must embrace interaction between communicative modes. The recent concepts of multiliteracy and multimodality are suggested for this purpose. This article examines why these concepts have emerged, the challenge they make to art education's traditional interest in the visual, and what teachers can begin to do in the classroom.

If we turn down the sound of a TV program, we find out how important dialogue is to our understanding of the picture. If we turn off the audiotrack to a video game, we find how critical the sound effects and music are to experiencing the game. And if we block out the written text on a website, we find that while the remaining images may be beautiful, or sensuous, or possess any number of qualities, it is usually impossible to say with any certainty what they are intended to mean. Each of these cultural sites involves more than the perceptual system of sight and more than the communicative mode of the visual. The same applies to magazines, shopping malls, theme parks, product packaging, advertising in all its forms, and all the other sites of contemporary visual cultural to which art educators are now turning their attention (as evidenced by recent art education journal special issues on visual culture) (Duncum, 2002; Sullivan, 2003; jagodzinski, 2003; Villeneuve, 2003). These sites use language as well as pictures, and many employ sound and music as well. Driven by global capital, these are the cultural sites whose signs now circulate day and night through international communication networks, saturate our cultural landscape, and impact our everyday thoughts and decisions. As Chapman (2000) says, "the culture of consumerism is not just visual" (p. 6).

The basis of this article is a clear and an obvious one: There are no exclusively visual sites. All cultural sites that involve imagery include various ratios of other communicative modes and many employ more than vision. While I will argue that this has always been the case with visual images, it is most evidently so with the cultural sites mentioned...
above. Furthermore, as Fairlough (2000) argues, the point of what he calls the “multisemiotic nature” of contemporary sites is not simply how one interprets discrete sign systems, but how meaning is extracted from how sign systems interact with each other (p. 162).

The consequence for art education of sign systems working together in the dominant cultural forms of our day is the need to rethink our traditional, exclusive focus on things visual. Having decided to study the visual culture of contemporary everyday life, it is necessary to acknowledge that no matter how important the visual characteristics of contemporary cultural sites are, in various ways and in varying degrees, they all involve other sign systems and appeal to multiple perceptual systems. The study of contemporary visual culture thus poses a major challenge to the basis on which we have long advocated and justified visual arts education, namely the exclusively visual nature of visual imagery.

In an attempt to unpack contemporary visual culture, to incorporate other communicative modes, I employ the recent term multiliteracy. Multiliteracy refers to two phenomena that are of concern to art educators (Cope & Kalantzis, 2000). First, there is the post-structuralist insight that any cultural site, of any kind, can be understood according to multiple readings generated from the multiple positions from which one views, reads, or hears. Secondly, all cultural sites, but especially ones like television and the Internet, include a range of modalities, especially language, images and sound. This second factor is referred to by the term multimodality. In this article, I focus mainly on the implications for art education of multimodality. However, I will not lose sight of the former meaning of multiliteracy because that would be to slip back into an older way of thinking that privileges texts over their contexts, and it is the social nature of literacy that is central to the enterprise of this article.

In defining literacy as primarily a social phenomenon, what matters are not correspondences between the specifics of communicative modes, but, as Buckingham (1993) says, understanding literacy as “utterance” (p. 28), which is “inevitably embedded within specific social practices” (p. 24). The term literacy is used to refer to the making of meaning with communicative modes. Multiliteracy is now employed to mean the making of meaning through the interaction of different communicative modes, or, as Cope and Kalantzis (2000) put it, “the multimodal relations between different meaning-making processes that are now so critical in media texts and the texts of electronic multimedia” (p.24).

Multiliteracy, then, does not involve just the abilities to interpret a picture, write a poem, calculate distance, or play a musical instrument. In each of these cases a different set of perceptual systems is involved, separate communicative modes are employed and separate literacies are enacted. I am concerned with how, in multiliteracy, different communicative modes interact with each other—how, for instance, words or

---

1 Cope and Kalantzis also use multiliteracy to refer to multiple versions of the English language such as American English and Asian English, but this meaning is not relevant here. Schwarzer (2001) uses multiliteracy to mean multilingual, in the sense that one can speak more than two languages, but again this is not relevant here.
even music without words can influence the meanings we attribute to a picture. For example, what needs to be said about a scene when the scene is shown in the form of a picture, and how do sound effects contribute to the meaning we make of pictures?

By contrast with this concern for interaction between communicative modes, teachers will be familiar with the use of one mode to inspire expression in another—for example, using a picture as stimulus for creative writing or a dramatic performance, or using music as stimulus for painting. This understanding also has a long history in academic literature on reading and writing in which educators have taken into account the visual or auditory dimensions to cognition as inspiration to the acts of reading and writing (Berthoff, 1984; Emig, 1983). Again, while one communication system inspires another, they remain separate. The definition of multiliteracy employed here requires interaction between two or more discrete sign systems. Literacy educators have also been aware of how in the act of reading and writing the mind works by simultaneously employing various sign systems (Berthoff, 1984; Emig, 1983). However, their interests were fixed on reading and writing, not on interpreting cultural forms that employed various sign systems.

The Need for Multiliteracy/Multimodal Education

For art education, the concern with multiliteracy and multimodality grows out of the current drive to reconceptualize the focus of art education as visual culture rather than art. The adoption of visual culture is part of a much broader movement within the sciences, social sciences, and humanities to conceptualize the visual as part of a general theory of communications (Mirzeoff, 1999). This movement arises out of a wholly new status given to the visual as a source of knowledge by many and disparate disciplines from astronomy to zoology (eg., Chaplin, 1994). It equally arises from an acknowledgment that in a wholly unprecedented way people today derive meaning from all kinds of imagery as part of their everyday experience.

For art education, the prime interest in the field of visual culture is not, of course, the scientific interest, but the cultural (Duncum, 2001). For us, the central concern with visual culture is the study of visual cultural sites/sights in terms of what they mean for personal and social life. Among visual culture scholars with whom we share this focus, the study of visual imagery is concerned with more than images; it is concerned with the whole context of images, their production and the lived experience of those who view and interpret (eg. Evans & Hall, 1999; Mirzeoff, 1999; Sturken & Cartwright, 2001). Studies in visual culture typically regard context as extending to the political and economic as well as the character of everyday life. Images are understood as constitutive of ongoing social practices, both a mirror and a contributor to social activity; and this focus is evident among art educators who have adopted a visual...
culture approach (eg., Duncum; 2002; Sullivan; 2003; jagodzinski, 2003; Villeneuve, 2003).

Sturken and Cartwright (2001) argue that studying visual culture means not only acknowledging the integration of the visual with other modes of representation, but how cultural forms inform life beyond their perception, how, in short, cultural forms are integrated into life. Others disagree and regard the term *visual culture* as problematic insofar as it foregrounds the visual and may appear to marginalize other communicative modes (Bolin & Blandy, 2003; Mirzeoff, 1999). The term *visual culture* appears to them to be made to do more than the term suggests, namely, to refer to communicative modes that are not visual. Whatever the merits of the term *visual culture*, the fact is that visual images appear in contexts that are invariably multimodal.

At the same time that art educators have begun to address contemporary cultural sites within their contexts, literacy educators have been moving to address how language is used in contemporary cultural sites (Anstey & Bull, 2000; Unsworth, 2001). Moreover, like art education’s focus on imagery in context, literacy education is increasingly viewing language as a social phenomenon. What were once minority interests within literacy education have now, like visual culture interests within art education, come to the fore.

Recognizing that literacy cannot be divorced from its social context has meant acknowledging that nowadays literacy almost invariably involves reading written texts in association with other sign systems. In the past the dominant view of literacy was that it was primarily a cognitive act, involving the mental processes of reading and writing. Literacy was considered a “largely fixed, individualistic and psychological ability” (Atkins, 2001, p. 11). Writing was understood often to be both inspired by and accompanied by imagery, but writing remained the focus (Berthoff, 1984; Emig, 1983). Similarly, reading was understood to be accompanied by visual imagery, but reading remained the focus. Being literate simply involved possessing the skills to decode and encode, and both skills were understood as mental operations. But now literacy is being understood to be a social practice (Slater, 1997). Instead of being located solely in the head, it is understood to be located in social settings, and, like images, instead of being located in texts themselves, it is understood to be located in contexts. An earlier semiotic approach that focused on texts alone has been replaced by an expanded social semiotic literacy that is grounded in social, including historical, contexts (Buckingham, 1993). Literacy is seen as dynamic. It is seen to change over time in response to changing applications of technology and social preoccupations, and, like visual imagery, to be profoundly political in the sense that it is used at every level with the intent to define and control the direction of events.
Thus, there now exists an emerging confluence between art educators concerned with visual culture and those in literacy education who are concerned with contemporary practices of language as social phenomena. Both fields come together in their acknowledgment that literacy has radically changed in both the world of work and the pursuit of leisure (Atkins, 2001). Physically intensive work has largely given way to the demand for a highly educated and flexible workforce that is skilled in using high-tech information systems, most of which are multimodal in nature. Similarly, in what Langman (2000, p. 3.) calls the "rise of the amusement society," leisure time has been greatly expanded, and much of it is spent enjoying cultural forms that are, again, multimodal. Increasingly, it is from these mediated cultural forms that people are deriving their reference points for living. As traditional institutions like the church, family, unions, and class affiliations wane in importance, people increasingly look to the cultural sites of global capital for their attitudes, values and beliefs (Jenks, 1996).

These observations about changing patterns of both work and leisure are the impetus for the study of visual culture. Yet, for the visual to be properly reconceptualized as part of a general theory of communications it must be understood in terms of how it operates as one sign system among others. The significantly new status that the visual has now gained should not prevent us from seeing it in dynamic relation to other sign systems. Among art educators, Marantz (1978) made this point long ago in relation to picture books. Consider, he suggests, the expectations that are set up by the color, texture and decoration of the cover, back page and endpapers of picture books as well as the expectations set up by one page for the next. As he says, written narratives and pictures obey their own conventions but "together they should provide a much fuller experience than either alone" (p. 83).

One of the most common uses of words and music in relation to pictures is to help anchor the meaning of pictures. A picture of a small child being led by the hand of a middle aged man towards some bushes with happy, bright music suggests a totally different meaning to the same image with sombre music. A picture of a child with a screwed up face with the words underneath, "Lost her doll" anchors a very different meaning than if the caption read, "Delighted at finding her doll." In both cases the preferred reading of these images is not contained within the image itself but in the relationship between image and music, and image and written text. And anchoring between image and text works both ways. Nodelman (1988) found that when he asked a range of people, children and adults, to retell a story from pictures in picture books where the words were obscured, they all expressed frustration and they developed a wide variety of interpretations. Here, as with the previous example, words were anchoring the meaning of the pictures. But when Nodelman read the texts without the pictures, his subjects often seriously
misunderstood the implications of the words. In this instance, the pictures were anchoring the meaning of the words. In both these examples, and all other multimodal sites, meaning lies within what we make of the space between discrete communicative systems.

The Uniqueness of the Visual

Considering communicative modes other than imagery goes against the long-standing basis of advocacy for visual arts education (Parsons, 1994). Since at least the late 1960s visual arts education has been justified and promoted on the basis of its uniquely, and exclusively visual nature. The meanings expressed by humans through visual forms have been conceived to have no literal equivalent in any other form of human communication. Both Arnheim (1969) and Goodman (1968) argued that thinking is domain specific, that thought can only be had in the particular form in which it is expressed. They proposed that it is impossible to transfer thought from one form to another without distortion. For them, visual and verbal thinking were equal but separate. For Eisner (1972), visual art offered a domain of human experience that was both a unique way of knowing and containing a unique content; it followed its own logic, possessed its own unique visual features, and, moreover, needed only to rely on its intrinsic visual qualities to be justified. Moving from one cultural form to another never involved mere transmission; it always involved translation, and translation always involved a creative process, which inevitably changed the original meaning. More recently, Gardner's (1983) theory of multiple intelligences has proved useful to art educators precisely because it provides theoretical support for justifying equal but separate emphasis on different ways of thinking and communicating.

There is, of course, no gainsaying the truth of the uniqueness of the visual at the level of a communicative code. Despite mistaken attempts by some semioticians to reduce the visual to language (see Buckingham, 1993, for a review), the visual does exist as a mode of both representation and communication that is independent of the verbal (Kress & van Leeuwen, 1996). Visual images exist as a relatively autonomous semiotic mode in which meanings are transported, made and remade with reference to no other mode but the visual (Unsworth, 2001). The visual deals with factors that language is patently ill equipped to handle, namely the visually salient elements of their subjects and of their spatial relations with each other (Kress, 2000). Ivins (1953) long ago demonstrated that the development of modern science, especially medicine, was only possible after the invention of the printing press which permitted exactly reproducible images, and today, from X rays to satellite surveillance, there is no substitute for visual imagery. Moreover, visual images are sensuous in ways that are particular to them. An emphasis on the uniqueness of the visual is well founded.
The Visual Was Never Exclusively Visual

However, to focus on the visual in this morphological way is to ignore the visual as a social practice. Art educators have focused exclusively on the visual nature of cultural sites only by excluding what is patently nonvisual in any actual use of visual material. While the impetus for multiliteracy education comes from recent developments in the workforce and entertainment industry, the traditional, exclusive focus of art education on the visual always involved a gloss. This is evident on examining both how children function in making and interpreting pictures and how adults, even within the professional artworld, make and interpret pictures.

Anstey and Bull (2000) point out that when children examine picture books they enjoy teasing out the relationships between pictures and words, and, the authors argue that this should not be surprising since even very young children's own early writing is often characterized by the creation of pictures to support their written narratives. The same is true of the unsolicited drawing of older children, when they use words to help carry the visual narrative or to anchor the meaning of images (Duncum, 1987). Children's own practices of both making and appraising images demonstrate the importance of the interaction between illustration and written text in making meaning.

That the uniqueness of the visual always involved a gloss on actual practice is equally evident from examining paradigms of modernist, abstract fine art, where the rationale for the work is said to lie with the image itself. Calling line, color, tone, and so on "elements of art," overlooked the fact that the cultural meaning of a particular painting or sculpture was always reliant on an interaction between people's prior knowledge, the artifact, and its title. The formal elements were the elements of the image, but not of meaning. Even in art galleries images are contextualized by language at both their immediate reception in the gallery space and in terms of the histories, criticism, and theories written about them. If we follow Foucault's (1981) notion of discourse, these works have no social significance outside the discourses in which they are created.

During the 19th century, the avant-garde rejected two of the previous functions of art, namely the recording of appearances and storytelling (Gowans, 1981). They turned, instead, to emphasize the visual qualities of art materials and the formal elements and principles of composition. Some of the nonrepresentational artists of the 20th century took this to its logical endpoint by claiming to eliminate all interests other than the image itself. Under the influence of phenomenology the viewer was meant, for example, to experience the Color Field painting and Minimalism of the 1960s purely in terms of the image (Archer, 2002). Every personal, social, political or economic consideration was to be put aside in favor of an exclusive focus on the work itself. Even the titles of
paintings were deliberately uninformative, some being called *untitled*. This represents the paradigm case of pure, undiluted imagery.

One could argue that the exception proves the rule, but even here the meaning of works was reliant upon language, for it is never possible to actually dissociate a painting from all its associations that are contained in language, and the deliberately uninformative titles acted like arrows to help convey the message that one was to focus on the imagery. By their very ubiquitousness, the titles directed viewers' attention and served to anchor the meaning of the images in the way titles normally act.

Parsons (1994) highlights the importance of titles by reference to a newspaper photograph that shows a dozen or so people in an art gallery, all of whom are looking not at the Malevich non-representational painting hanging on the wall, but at the accompanying text. Even when there is the most minimal information provided—the artist's name, the title of the work, its media and its date—meaning is provided in part through language, or rather through what people make of the interaction between the language used and the artifact. Furthermore, while words often appear in some older artworks (such as paintings from the early Renaissance) today, seemingly in flight from the dominance of pictures elsewhere, many artists are now using words as a substantive part of their work, and some use words exclusively (Archer, 2002). Many artworks from the past that did not rely on words within images relied on references that were known through language such as biblical or classical tales. A focus on the exclusively visual to justify art education was, then, always specific to a limited range of historical developments, and even then it was flawed. Pictures, even non-representational paintings and sculptures, were always multimodal in terms of social practice.

This theoretical point having been made, it is clear that there is no avoiding the multimodal nature of dominant and emerging cultural sites. Whether it is television, the Internet, zines, video games or simulation rides, each is clearly a hybrid of communicative modes. Moreover, one does not read the language and then the pictures and then listen to the sounds; rather, one takes them in as a *gestalt*, a whole, all at once. This then is the challenge of multimodality for education.

**Multimodality and its Pedagogy**

What follows here is only provisional, for we have hardly begun to deal adequately with the multimodal nature of communication. Yet there appear to be at least two different approaches that seem profitable which have been advanced as media education and multiliteracy education.

Media education (eg. Lacey, 1998) focuses largely on the interaction of spoken and written text and images, though it also deals with music and sound effects. It primarily focuses upon print media, mostly advertising, film and television. In studying film, the principal teaching tools of media education are to study segments of film, and for students to storyboard
and make their own films and videos. In this regard it is similar to art education, which also asks students to study visual forms and, being informed by such study, to make similar forms. In studying film, a common method is to examine segments of film, repeatedly going over the same short segment, sometimes without the audio track, sometimes with just the audio track. This is necessary because students are so familiar with the whole of the experience that they are unaccustomed to attending to its separate parts. Students are asked to consider how information and mood are conveyed: is it through camera angles; point of view; length of shots; framing or cropping devices like long shot and close up; editing techniques like fade in/fade out, dissolves or wipes; dialogue; ambient sounds such as sound effects created on screen; or sounds which originate off screen such as a voice-over or music? Students examine nonverbal communication such as the facial expression and body language of the actors, their gaze, gestures, posture, body contact, and clothes. They examine what aspects dominate, and how the sounds sometimes mirror the image and sometimes work against it. For example, I have used a 30-second segment from the film *Woman on Top* that shows a man walking along a foreshore and coming up and engaging in a brief conversation with fishermen tending to their nets (Case & Torres, 2000). By watching this brief segment over and over again, students became aware of how the syncopated beat of the Latin American music connects to the movement of the waves in the background. They recognized how the wind plays upon the actors’ hair and loose-fitting clothes, all echoing the central figure’s sense of unease. Students saw how the relatively long opening shot set up and contrasted with the short cuts of the brief exchange between the men. In making their own films, students must consider all these factors: how to light a scene, how to edit, what music to use and at what pitch, what dialogue is necessary, what sound effects to use, and so on.

Multiliteracy education has been frequently concerned with the relationship between written words and images in children’s picture books. Some art educators have already worked with this cultural site (eg. Cross, 1996; Eubanks, 1999; Marantz, 1978), and theirs is an excellent choice. Anstey and Bull (2000) suggest that postmodern picture books, like Anthony Browne’s *Gorilla* and *Voices in the Park*, are among the most innovative sites for exploring complex relationships between pictures and words. These texts are characterized, in part, by “an ironic or unusual relationship between written and illustrative text” (p. 189). Nodelman (1988) comments that placing words and pictures together “invariably changes the meaning of both so that good picture books as a whole are a richer experience than just the simple sum of their parts” (p. 199).

Picture books are not only for young children. In Australia and the United Kingdom such books have long been used with senior high school students.
students as part of the English syllabus to help facilitate an understanding of the constructed nature of both visual and verbal texts. Students examine the "subtle interweaving of words and pictures, varieties of meanings suggested but never stated, visual and verbal clues to intricate patterns, structures and ideas" (Chambers, 1985 cited in Unsworth, 2001, pp. 261-262). Literacy educators have tended in the past to see the illustrations as merely mirrors of the written text but it is crucial, as art educator Marantz argued long ago, to examine the functional relationships between words and pictures in a much more complex way. Questions to ask of a functional kind include:

- how aspects of characterisation, plot, settings and theme were carried by the illustrative and written text. Were they shared equally, did one text take a more dominant role at times and if so why?
- What combination of elements in the balance and layout portrayed these aspects in the written text and how were these aspects portrayed in the illustrative text? (Anstey & Bull, 2000, p. 188)

Often it is a matter, as Nodelman (1988) puts it, of pictures showing what the words do not tell and the words telling what the pictures do not show, or, as Marantz (1978) says, often words suggest while pictures stipulate, although in some recent picture books these roles are reversed.

In the classroom, Unsworth (2001) has students study a range of picture books on the same story or theme such as bullying or the supernatural. Alternatively, he adopts a strategy suggested by Marantz (1978). He has students study across genres, including picture books, novels, and illustrated storybooks. In this way, the nuances of interpretation in any one source become more readily available than if studied separately, and so do the relationships between image and text as employed in different books. Cross (1996) explored this approach with picture books that told very different versions of the story of Christopher Columbus landing in North America. She compared different approaches to Columbus: the hero and the dreamer. She found that books offering Columbus as a hero used a biographic written format and thereby suggested they were documenting true historical facts. These books used a variety of distancing devices such as having considerable white space around both images and text, often using frames around both, and separate pages for text and image. Invariably, a realistic style of image was employed, and there was a focus on maps and navigational tools. Typically, such books were small and narrow and used glossy paper stock and, by these various devices, complemented the written text as offering an objective, unemotional view of events. By contrast, books suggesting that Columbus was incompetent but could be forgiven for being a dreamer, used quite different visual cues. They tended to be larger in format, often used a cartoon or painterly style of illustration, often used saturated colors, and used white space to draw the reader into the emotional quality of the image, not to distance...
the reader. The tone was warm and inviting. Cross also reviewed picture books that critique both traditional views of Columbus and found conflicts between the images and written text that allow readers to construct their own interpretations.

Of course these are only starting points. We must look beyond existing practice to consider the relationship of words, sounds, music, and pictures in a wide range of cultural sites, including theme parks and department stores, tourist sites and food outlets, music videos and the Internet. Nevertheless, current classroom practice suggests places for us as art educators to start to deal with the sites of corporate, global capital in terms of multimodality and meaning.

**Conclusion**

In 1934, Walter Benjamin predicted that there would be an increasing inter-dependency of words and images, and he called for a citizenry equipped with a critical faculty with visual and verbal communication. His was an urgent call because at the time the Nazis were advancing the synergetic use of every available mass medium (Evans & Hall, 1999). The captains of global culture are not as immediately dangerous as the Nazis, but they do have available to them a much greater array of multimodal media than was ever previously available, and their primary interest lies in profits, not the public good. The cultural forms of global capital combine images, words, and sound to produce highly seductive experiences that are not in everyone's best interests. The need for a citizenry equipped to deal with multimodal cultural sites remains pressing. It is greater than any rationale that might be mounted for visual arts alone, or language alone, or music alone, or perhaps any other school subject.

**References**


Visual Culture Isn’t Just Visual: Multiliteracy, Multimodality and Meaning


