Cognitive Effects of Multimedia Learning

Robert Z. Zheng
University of Utah, USA
Chapter V
Theoretical and Instructional Aspects of Learning with Visualizations

Katharina Scheiter
University of Tuebingen, Germany

Eric Wiebe
North Carolina State University, USA

Jana Holsanova
Lund University, Sweden

ABSTRACT
Multimedia environments consist of verbal and visual representations that, if appropriately processed, allow for the construction of an integrated mental model of the content. Whereas much is known on how students learn from verbal representations, there are fewer insights regarding the processing of visual information, alone or in conjunction with text. This chapter uses a semiotics approach to provide a definition of visualizations as a specific form of external representation, and then discusses the differences between verbal and visual representations in how they represent information. Finally, it discusses how meaning is achieved when learning with them. The next section discusses basic perceptual and cognitive processes relevant to learning with visualizations. This background is used to specify the instructional functions that visualizations have either as self-contained instructional messages or as text adjuncts. Moreover, the role of individual differences in processing visualizations is highlighted. The chapter ends with methodological suggestions concerning the important role of interdisciplinary research and assessment methods in this area.

Copyright © 2009, IGI Global, distributing in print or electronic forms without written permission of IGI Global is prohibited.
INTRODUCTION

Visualizations constitute a key component in multimedia-based instruction, which can be defined as learning from text and pictures (e.g., Mayer, 2005). Despite the fact that visualizations are used more and more frequently in informal and formal educational settings, not much is understood about their semiotic properties, how humans process them, and how they can be best designed to learn from. In educational research, visualizations are often treated in a uniform manner, despite the fact that the visualizations might serve vastly different functions depending on the audience and goals. Just as bad, visualizations are treated as functionally equivalent to text. As a consequence, reviews on learning with visualizations are equivocal, with studies showing widely varying effects (negative to positive) on learning. In the current chapter, we will try to provide a more differentiated view by first reviewing the literature from different disciplinary perspectives (education, semiotics, perception, and cognition) to characterize different types of visualizations, to distinguish them from verbal representations, and to describe how information is derived from them. This approach will attempt to provide a unique approach to addressing the question of when and why visualizations are effective for learning. After some summarizing remarks, directions for future research will be outlined in the final section of this chapter. It is important to note, however, that we will not review the more mainstream literature on the effectiveness of learning with visualizations, as comprehensive reviews can be found elsewhere (e.g., Anglin, Vaez, & Cunningham, 2004; Rieber, 1994).

BACKGROUND

What are Visualizations?

Visualizations are a specific form of external representation that are intended to communicate information by using a visuo-spatial layout of this information and that are processed in the visual sensory system. According to Rieber (1990, p. 45) “visualization is defined as representations of information consisting of spatial, nonarbitrary (i.e. “picture-like” qualities resembling actual objects or events), and continuous (i.e., an “all-in-oneness” quality) characteristics”. Visualizations are often best understood through the context of their use (MacEachren & Kraak, 1997). In as much as visualizations are created to communicate in a learning or problem-solving context, these visualizations are typically based on models and leverage human perceptual and cognitive abilities to efficiently and effectively convey information (Gilbert, 2005). The model and its use in the context of the visualization then drive particulars of the visualization—from the visual metaphors employed to the dynamic characteristics of elements (Bertoline & Wiebe, 2003).

External representations such as visualizations are defined with regard to their relation to the real world. “The nature of representation is that there exists a correspondence (mapping) from objects in the represented world to objects in the representing world such that at least some relations in the represented world are structurally preserved in the representing world” (Palmer, 1978, p. 266). Thus, a representation is defined through its structural correspondence to what it stands for (i.e., the referent) and is hence analogical to the referent. By means of this analogy, representations can act as a substitute for the referent and evoke similar responses as the real-world referent. Semiotics is an approach that can be used to more rigorously analyze the relationship between the signs that make up a visualization, the underlying intended instructional message of the visualization, and the learning task context in which the visualizations are being employed.

Using semiotics as a methodology, visualizations can be understood and organized in ways that better guides their creation and intended use. Peirce (1960) identified three forms of relations between the representation and the represented...
object: an *icon* resembles the object it depicts in terms of its criterial attributes in a given context. Criterial attributes are properties of the object that "act as discriminanda for sorting and resorting the objects in the perceptual world" (Knowlton, 1966, p. 162). An *index* refers to its object by means of a physical connection to it (e.g., a footprint as a representation of a bear). A *symbol* (or digital sign, according to Knowlton, 1966) bears no resemblance to what it stands for and is thus arbitrary (e.g., words, numbers). Knowlton (1966) further distinguishes between realistic pictures, analogical pictures, and logical pictures.

**Realistic pictures** resemble their referents by means of physical similarity (e.g., a picture of a tree looks like a tree we see with our eyes). This similarity is achieved by copying the real-world referent with respect to shape, details, color, or motion. From a semiotics perspective, there are at least three problems associated with the labeling of this category of visualizations. First, constructivists like Gombrich (1969) have proposed that what people perceive as being realistic in a picture is strongly affected by their preconceptions of how a representation of the real-world object should look like. Second, realistic pictures may vary largely with regard to the level of realistic detail, which is why Alessandrini (1987) suggested calling this type of pictures 'representational' rather than 'realistic'. The third problem with this labeling is that it does not help distinguishing the first from the second visualization category, namely analogical pictures.

**Analogical pictures** may seem realistic in the sense that the depicted objects look alike to something known from the phenomenal world. However, analogical pictures “make reference to something else – something that is in some way analogous to the portrayed object or to its manner of functioning” (Knowlton, 1966, p. 176). Thus, the objects depicted in the visualization refer to entities that are different from entities that they resemble (e.g., two lumberjacks moving a felled tree as an analogous picture for muscles moving a bone). Resemblance to the referent is established solely through functional similarity or structural correspondence. Whereas realistic pictures can be used to represent the phenomenal world only, “analogical pictures can represent either the phenomenal or nonphenomenal world. In both cases, this is done through the bridge of the ‘visual’ phenomenal world” (Knowlton, 1966, p. 177). Accordingly, these pictures can represent abstract concepts, albeit not directly; rather, “they can be illustrated indirectly by showing their effects on visible objects, tangible results, specific instances, or concrete exemplars” (Alessandrini, 1987, p. 176).

**Logical pictures** can be thought of as highly schematized pictures, where the elements of the picture do not bear any physical resemblance to objects in the phenomenal world and are thus arbitrary (cf. arbitrary pictures, Alessandrini, 1987; Rieber, 1994). Accordingly, one might argue that logical pictures are symbols rather than icons. On the other hand, logical pictures, like analogical pictures, preserve the structural interrelations between these referents. Examples of logical pictures are depictions of things that are potentially visible, but where we do not know how they look like (e.g., atoms) or of things that do not exist in any tangible way (e.g., a mathematical proof). Moreover, the term logical (or arbitrary) picture is used to describe charts, diagrams, and graphs, because here the spatial layout is used to convey information on conceptual relationships. Logical may be the better term in this case since conventions for the design and layout of charts, diagrams, and graphs are often far from what would typically be considered arbitrary (Bertin, 1983).

Semiotic taxonomies like the one by Knowlton (1966) have been largely neglected in educational research (Anglin et al., 2004), although they may be very relevant to the research and development of visualizations. Research on learning with visualizations has often made generalizations across a wide range of types, which have led
to the misuse of multimedia principles (Mayer, 2005). Moreover, visualizations have been incorporated in instruction with the assumption that they behave the same as verbal representations. While both text and graphics can be scrutinized through semiotic analysis, they are truly unique sign systems with differing strengths and weaknesses in communicating referents.

How Do Visual Representations Differ from Verbal Representations?

The properties that make visualizations different from verbal representations have been discussed in various disciplines. Differences between the two representational formats pertain to three aspects: (1) the features of the external representation, (2) the external representation’s relation to the represented world, and (3) the relation between the external and the internal representation (i.e., the representation’s meaning).

Features of the Representational Formats

Modularity. Kosslyn (1994) has suggested that verbal and pictorial representations differ with regard to their modularity. Verbal representations can be broken down in small, discrete symbols (i.e., letters), however breaking down letters into even smaller parts will not yield meaningful pieces. On the other hand, there is no discrete unit for visualizations, because they can be arbitrarily broken down in multiple ways and still yield potentially meaningful patterns.

Sequentiality. Visualizations comprise spatial relations to represent information in a two- or three-dimensional structure. Linguistic information, on the other hand, is presented in a purely sequential manner (Larkin & Simon, 1987).

Modality. Verbal representations are amodal, in that information extraction is not linked to a specific perceptual modality. That is, verbal information can be processed by the auditory system, the visual system, or even by the haptic system. On the other hand, the processing of visual representations is strongly associated with the visual modality.

The Relation between the External Representation and the Represented World

Arbitrariness. Verbal representations are said to be arbitrary in the sense that there is no inherent reason for why the word “house” denotes the thing we live in, for instance. Rather, the object could be represented by any given arrangement of letters or phonems if a culture agreed on this. On the other hand, most visualizations are non-arbitrary in that the representation and the represented world are linked to each other by means of inherent structural correspondences or even by physical resemblance.

Notationality. In verbal representations, distinct elements are linked to real-world elements in an unambiguous and explicit way, thereby determining the semantics of the representational elements (Goodman, 1976). Moreover, for every language there are explicit rules telling us how to combine words to sentences or how to express temporal relationships, thereby determining the syntax of the representation. Goodman calls this the notationality of a representational system, which denotes the “degree to which the elements of a symbol system are distinct and are combined according to precise rules” (Anglin et al., 2004, p. 870). Visual representations are far less explicit in terms of their semantics, whereby these elements can be represented in many different ways and still be recognized, as long as the structural relations are preserved. Nevertheless, in semiotic theory there have been attempts to determine the regularities of visualizations and, therefore, to identify their underlying syntax (Kress & van Leeuwen, 1996). Logical pictures may lend themselves best to this approach since they are often notational in that there are distinct elements arranged according to specific rules or conventions (Tversky, 2003).
Parsimony. The famous saying “a picture may be sometimes worth 10,000 words” suggests that it sometimes takes 10,000 words to express what can be depicted in a picture with markedly fewer symbols. One reason for this parsimony is that spatial relations of objects are automatically implied in a picture and thus do not require any additional symbols, whereas in verbal representations spatial arrangements need to be made explicit (e.g., $x$ is positioned above and to the right of $y$). According to Kosslyn (1994) visualizations are more complete than verbal representations in that they deliver information on all possible details of the object (e.g., its shape, size etc.) that is being represented through them.

Expressiveness. Contrary to verbal expressions, visualizations have limited power for expressing logical and temporal relations as well as abstractions (Maybury, 1995). For instance, to represent the concept “animal” it would not be very helpful to show a dog or a cat, as these would be interpreted as standing for the concrete animal, but not for the superordinate category. Knowlton (1966), however, notes “that it may be possible to suggest a concept with an iconic sign, depending, in part, on its level of realism” (p. 167). Accordingly, Schwartz (1995) demonstrated that realistic visualizations made people think more about the concrete referent of the depiction. On the other hand, schematized visualizations facilitated abstract reasoning and symbolic interpretations of the represented objects (DeLoache, 1995; Goldstone & Son, 2005; Schwartz, 1995).

Specificity. According to Stenning and Oberlander (1994) and Bernsen (1994), visualizations can be characterized by their higher specificity compared to verbal representations. That is, while language can be interpreted in multiple ways, thereby corresponding to multiple possible represented worlds, visualizations are often more constrained in that respect and reduce the number of possible worlds that can be represented through them (cf. graphical constraining, Scaife & Rogers, 1996).

The Relationship between Internal and External Representations

Conventionalism. The meaning of a verbal representation (i.e., what it stands for or what it signifies) is established through convention or cultural agreement. This is why Palmer (1978) referred to verbal representations as extrinsic representations, where meaning is imposed onto the representation from the outside. The meaning of the visualization is intrinsic to the representation as it is constructed based on the properties of the objects it represents; thus, its meaning is derived from these inherent characteristics. Whereas these structural correspondences can potentially be expressed in very many ways, for some visualization forms there are established cultural rules as to how to represent the structural features of the real world, thereby constraining the multitude of options (cf. Gombrich, 1969; Kosslyn, 1994). As outlined before, this is particularly true for logical pictures (Tversky, 2003).

Interpretation and Reasoning. When being confronted with a verbal expression for the first time (e.g., as a child or as a second-language learner) we will not understand this expression unless we are explicitly told its meaning. Realistic visualizations, on the other hand, can be intuitively understood without making the link to the real-world referent explicit. This is even true for babies who lack experience with external representations (DeLoache, 1995; Hochberg & Brooks, 1962). Gibson (1979) has argued that this is because realistic visualizations result in optical arrays similar to those when looking at their real-world referents. Hence, information “can be directly picked up without the mediation of memory, inference, deliberation, or any other mental processes that involve internal representation” (Zhang, 1997, p. 181). Similarly, other researchers have suggested that visualizations may support inference and reasoning processes grounded in perception (Goldstone & Son, 2005; Schwartz, 1995; Stenning & Oberlander, 1994).
allowing perceptual judgments to substitute for more demanding logical inferences. While these arguments are evident for realistic or representational visualizations (Alessandrin, 1987; Knowlton, 1966), logical pictures may also have specific cognitive processing benefits compared to verbal representations. Larkin and Simon (1987) have suggested that diagrammatic visualizations are often more computationally efficient for accomplishing tasks that require the processing of visuo-spatial properties (cf. cognitive offloading, Scaife & Rogers, 1996). In particular, visualizations reduce the need of searching for multiple information elements related to a single idea because this information is grouped (chunked) in visualizations. Thus, in visualizations related information can be processed in parallel (to a certain extent), rather than sequentially.

**Internal representation.** Verbal and visual representations differ in their dominant internal representation formed in long-term memory. Whereas verbal descriptions are associated with the construction of a propositional representation, non-verbal formats like visualizations are more likely to be encoded and stored as analogical representations (Kosslyn, 1994; Paivio, 1991). According to Paivio, these two internal representations are interconnected by referential links so that they can activate each other (e.g., the word “apple” activating the mental image of an apple tree). It is important to note that words and visualizations may, in principle, both yield a propositional representation as well as an analogical representation (e.g., through construction of a mental image for concrete words). The so called picture-superiority effect (i.e., better recall for pictorial than verbal information) is explained by assuming that this dual coding of information based on a single input representation is more likely to occur for pictures than for words. The picture-superiority effect can also be explained by Baggett’s *bushiness hypothesis* (e.g., Baggett, 1984). It states that knowledge acquired from visual rather than verbal external representations will be better accessible in memory because the respective nodes in memory share more associations with other nodes in the semantic network. Thus, visual concepts are assumed to be “bushier” than verbal concepts and, therefore, more salient in memory.

Bernsen (1994) mentions a number of mechanisms that help to overcome inherent weaknesses of the two representational formats, such as focusing mechanisms in graphics and specificity mechanisms in language. Thanks to focusing mechanisms (e.g., selective removal of specificity, highlighting, selective enlargement), analogous visualizations come closer to the strengths of a natural language. Thanks to specificity mechanisms in language (summaries, key points, usage of metaphor and analogy), language comes closer to the strengths of an analogous representation. Even with these mechanisms being present, visual and verbal representations will differ on a variety of dimensions, which naturally affects the way they are processed during multimedia learning. While there are elaborated models of text comprehension, no comprehensive models exist for the processing of complex visual information. In the following section, some pivotal concepts of visual perception are described as they are relevant to multimedia instruction.

**Processing of Visualizations**

While the research traditions in understanding text and graphic representations are equally rich, visual perception has been less well represented in educational research. Understanding the perceptual and cognitive processing of graphics is crucial to understanding the differences between text and graphic representations and how to most effectively design visualizations.

The processing of visualizations for learning and problem-solving can be understood in the context of information processing theory and, more specifically, multiple resource theory (Wickens, 2002). Using this approach, the processing
Theoretical and Instructional Aspects of Learning with Visualizations

of visual images contains both perceptual and cognitive elements, each of which require their own resources. In addition, there will also be resources dedicated to overarching, executive functions, guided by both short-term and long-term goals (i.e., the task). While earlier information processing models were often depicted as consisting of a number of distinct stages and functions, multiple resource theory recognizes that mental functions differ both in terms of the amount of resources required and the degree of resource overlap with other functions. Common to information processing theory, resources related to both sensory buffers and short term memory are finite while long term memory resources can be considered essentially infinite. It is the resource-limiting nature of sensory and short term memory and the ability (or lack thereof) to access relevant long term memory resources that informs much of the research in multimedia instructional design (Mayer, 2005; Sweller, van Merriënboer, & Paas, 1998).

It is useful to think of perception unique from cognition (Hochberg, 1978; Zhang, 1997). Perception can be operationally defined as those (visual) mental functions that do not require access to long-term memory. These, then, are abilities common to the normal adult and adolescent population and can be assumed when designing visual representations. Perception, therefore, involves basic processes such as discrimination based on visual properties of color, shape, texture, and orientation, form recognition and spatial arrangement.

Within perception, there is both ambient and focal vision, with the difference largely based on attention (Wickens, 2002). In terms of resources, attention is an executive function that allocates resources to processing information received by the sensory buffers. Some perceptual functions—including peripheral vision—make use of low-level processing capabilities in the brain that require little or no resources (i.e., pre-attentive or automatic processes). These ambient perceptual capabilities include supporting much of everyday functions such as basic hand-eye motor skills or locomotion through space. All perceptual processes are actively involved in sense-making. These low level ambient functions will attempt to organize a scene as three-dimensional space, segregating foreground and background elements and determining whether these elements lie in discrete locations in depth (e.g., a planar surface) or as a continuous three-dimensional form. Elements will be grouped based on visual properties and their spatial relation to each other. While some of these properties, such as color, are pre-attentive, others (e.g., shape) may require additional attentive processing (Treisman & Gelade, 1980). The Gestalt principles of configuration are some the best known examples of active perceptual organization of a scene (Rock & Palmer, 1990). These ambient processes can be seen at work initially organizing visual information used in multimedia instruction.

Focal vision requires attention and thus more resources than ambient vision. Focal vision will make use of both bottom-up information gathered from ambient vision and top-down information provided by both ongoing focal viewing and higher level cognitive processing. Ambient visual function may provide the “gist” of a scene whereby focal vision then attends to elements initially deemed as interesting, unusual or important. Ongoing processing of this scene then guides focused vision. In addition to information being passed from ambient vision to focused vision functions, information processed at this next level may be processed further and passed to long term memory. Eye movement research informs our understanding of this ongoing process (cf. Henderson, 1992). While our conscious perception of a scene is of a single, unified, stable image, the underlying eye movements reveal constant saccadic movement, sampling various portions of the scene to varying degrees. The brain, employing both ambient and focal vision, engages in both real-time, transient and longer-term meaning making of what is received in the sensory buffers and creates the sense of a stable, meaningful scene.
Cognition, in this context, can be thought of those resource-driven processes that require long-term memory. For example, viewing a photograph of a lion, perception would be able to separate the lion from the surrounding background, use shading and perspective convergence to mentally orient it’s three-dimensional shape in space. Prior knowledge would, however, be needed to know that this form was a lion. Further knowledge would be needed to know whether this lion was male or female, and still further specialized knowledge might allow you to use visual cues to discern the approximate age of the lion. In the context of learning, the difference between perceptual and cognitive functions becomes very important. While most perceptual capabilities can be assumed to be universally present in a population, cognitive capabilities may be much more varied (Mayer & Massa, 2003; Siegler & Alibali, 2005). While all of the students in a sixth grade classroom may know that the photograph is of a lion, they may not all know its sex or approximate age. What is important is which of these elements of prior knowledge is important for the learning task and/or whether these elements are what you hope the student will learn about. Prior knowledge is equally important with analogical or logical images, where the visual sign element must be mapped to its referent in memory. Improper scaffolding may leave a student at a lower level of processing, where the image seen is just a “red triangle” and not, say, a representation of a key biological molecule used in DNA replication (Patrick, Carter, & Wiebe, 2006).

The movement of visual information from sensory buffers to short term memory and then possibly to long term memory establishes a number of points of possible failure to encode into schemas information necessary for a learning task. Low level visual information that cannot be apprehended because it is too small or does not possess enough contrast to be segregated from surrounding elements cannot be used as part of higher level visual processing. Similarly, elements that meet the requirements of low level apprehension may still not be further processed because attention has not directed the necessary resources to this visual information; the information is not rehearsed in short term memory and not integrated into long-term memory. Finally, visual information in long term memory may not be retrieved because it is not recognized as related to the current visual processing task.

It is important not to think of these processes as rigid and linear, as there is an ongoing flow of information between all levels in both directions, with some of it happening in parallel while other elements happening in a more serial fashion. In addition, it is equally important to not make assumptions about the structure of visual information as it is stored or processed. There is no reason to believe that there is a direct, homologous relationship between the distal visualization (whether it be a physical three-dimensional object, printed graphic, or computer display) and the internally coded visual information (Scaife & Rogers, 1996). That is, that one’s internal mental images, no matter how they might “look” like their corresponding real world objects, can be composed of information from multiple sources, including both real-time sensory streams and long-term memory traces of previous experiences. Ongoing cognitive and neuroscience research is continuing to unravel both the physiological and psychological basis for visual information processing and storage.

LEARNING WITH VISUALIZATIONS

If visualizations are going to be optimally deployed in instructional materials, their semiotic, perceptual, and cognitive characteristics and their relationship to textual elements need to be synthesized and operationalized. In the following sections, two issues will be addressed that are relevant to learning with visualizations. The first section aims at specifying the conditions under
which visualizations will contribute to learning. The second section addresses the role of learner characteristics in visualization research.

When Do Visualizations Aid Learning?

A main issue in designing instructional materials pertains to the questions of when and how to use visualizations for conveying knowledge in a particular domain. To find an answer to these questions, one has to ask what information is to be conveyed, what the users’ physical and cognitive abilities, preferences, and intentions are, and what the communicative situation is where the instruction will be used (Maybury, 1995). In the following section, we will highlight the function a visualization may have for learning and whether this function is relevant to achieving a particular learning objective. Hence, visualizations are supposed to be effective for learning only if they fulfill functions relevant to achieving a particular learning objective. In the following, we will distinguish instructional functions that visualizations may have if presented as self-contained messages (i.e., regardless of verbal information) or if they accompany verbal explanations. We will end this section with a discussion on whether visualizations can aid learning even if they are redundant to other representations.

Instructional Functions of Visualizations as Self-Contained Messages

The design of the visualization, of course, has to take the instructional context into consideration. There are several instructional functions that visualizations may have—irrespective of whether they are accompanied by verbal explanations or not.

Affect. Visualizations are often said to be motivating for students, because they can make a subject matter more interesting and appealing to students. Moreover, they can trigger specific emotions or lead to a change in learners’ attitudes (Levie & Lentz, 1982). However, there is a danger that visualizations that do not convey any information relevant to the learning objective will distract learners (cf. seductive details effect, Harp & Mayer, 1998).

Replace and augment real-world experience. Visualizations as instructional materials can act as a substitute to direct experience by offering highly realistic impressions of real-world objects and events, which might otherwise be too small, too large, too fast, too far away, or too dangerous to observe in reality. In that respect, visualizations do not replace real-world experience, but they may even improve this experience by providing information that would not have been accessible in the real world (i.e., “better than reality”).

Visuo-spatial reasoning. As has been outlined in the section on differences between verbal and visual representations the latter provide direct and parsimonious access to visuo-spatial information and, in case of dynamic visualizations, temporal properties of objects and events (Larkin & Simon, 1987; Rieber, 1990; Tversky, Bauer Morrison, & Betrancourt, 2002). This visuo-spatial information can also be directly used for inferences and reasoning (Goldstone & Son, 2005; Schwartz, 1995). With verbal representations, the sequential information would not only require cognitively demanding processes of searching related information, but also it would need to be internally transformed into a form more appropriate for spatial reasoning. Thus, for visuo-spatial reasoning tasks visualizations are more computationally efficient and allow for cognitive offloading (Larkin & Simon, 1987; Scaife & Rogers, 1996). This function is important in that it does not limit a visualization’s capabilities to act as a memory aid; rather, it suggests that reasoning based on visualizations is a way of extending the boundaries of cognition to external representations (Zhang, 1997). External visualizations enable cognitive operations that would otherwise have to be conducted internally (e.g., mental imagery)
and thereby require more cognitive effort (cf. supplantation, Salomon, 1979).

**Instructional Functions of Visualizations as Text Adjuncts**

Most often, visualizations are used in conjunction with verbal instructional materials rather than acting as self-contained messages. Analyzing the combination of text and pictures has a long tradition within semiotics and sociosemiotics (Eco, 1976; Kress & van Leeuwen, 1996), in interface design (Bernsen, 1994; Mullet & Sano, 1995) and in human–computer interaction (Maybury, 1995) and textlinguistics. The perspectives in these works differ; while some stress the differences between text and pictures with regard to their suitability for expressing specific information (Bernsen, 1994; Mullet & Sano 1995), others are more interested in the interplay between them. The latter perspective is also taken here in this section, where it is emphasized that when visualizations accompany text they have functions in addition to those mentioned before, because they interact with the verbal information in specific ways. The most prominent analysis of the instructional functions associated with such a use of visualizations as text-adjuncts in the education literature has been conducted by Levin, Anglin, and Carney (1987). In their review, the authors described five functions of visualizations as text adjuncts: decorative, representation, organization, interpretation, and transformation.

Visualizations with a *decoration function* are not related to the verbal information and are introduced only to make a text more appealing and interesting for learners. However, while they are supposed to motivate learners, the meta-analysis by Levin et al. (1987) actually revealed a negative effect of decorative visualizations (cf. seductive details effect, Harp & Mayer, 1998). Presenting irrelevant additional information may distract learners from processing the pivotal learning contents or it may trigger inappropriate schemas for encoding the relevant content.

Visualizations in their *representational function* depict objects and relations mentioned in a text in a way that the meaning of the text more accessible for learners by making a text more concrete. Visualizations with an *organization function* provide an organizational framework for a text (e.g., how-to-do-it diagrams) and thereby make the content more coherent by highlighting the argumentative or organizational structure of the text (Verdi, Kulhavy, Stock, Rittschof, & Johnson, 1996). According to Levin et al. (1987) this function refers to comprehensible texts, whereas the *interpretation function* of visualizations is to make texts understandable for learners that would otherwise be incomprehensible. Accordingly, these visualizations are often introduced in textbooks and multimedia instructions to clarify difficult-to-understand passages and abstract concepts within passages (e.g., pictorial analogies).

The rarest function that visualizations are used for pertains to the *transformation function*. Visualizations that follow this function are designed to improve memory performance directly “by targeting the critical information to be learned, and then (a) *recoding* it into a more concrete and memorable form, (b) *relating* in a well-organized content the separate pieces for that information and (c) *providing* the student with a systematic means of *retrieving* the critical information when later asked for it.” (Levin et al., 1987, p. 61). Visualizations with a transformation function showed the strongest positive effects on learning outcomes in the meta-analysis conducted by the authors.

It is important to note that Levin’s review is limited to prose learning, whereas most multimedia materials are concerned with a broader and deeper understanding that can be transferred to novel situations. It is thus questionable whether these recommendations can be simply applied to the design or multimedia materials. There is at least one recommendation where research seems to disagree: informational equivalence of verbal and visual representations. According to Levin et
al. (1987), students who use redundant information to acquire content knowledge benefit from visualizations if these show a large information overlap with the verbally presented information. In fact, the representational, organizational, and interpretational function of visualizations presuppose that text and picture are at least partially redundant. On the other hand, multimedia learning research, has suggested not to present redundant information to learners, because this may require additional cognitive resources for comparing and integrating the information, which are then no longer available for learning (redundancy effect; e.g., Bobis, Sweller, & Cooper, 1993).

There are at least three things that are usually ignored in this discussion in the literature on multimedia learning: First, a certain degree of overlap is necessary to allow for a coherent mental representation, where learners can draw connections between the two sources of information. Second, what is redundant information is often impossible to say. Even when deliberately trying to construct informationally equivalent (and thus fully redundant) representations, it is almost impossible to counteract the fact that visualizations often “unintentionally” convey more, albeit potentially irrelevant information than do verbal descriptions due to their higher specificity or completeness (Kosslyn, 1994; Stenning & Oberlander, 1994). Third, as has been emphasized by Ainsworth (1999, 2006) in her functional taxonomy of multiple external representations (MERs), even informationally equivalent representations may have different functional roles for learning and thus support knowledge acquisition differently. She categorizes these roles into three groups: First, visual and verbal representations may fulfill complementary roles in instruction by facilitating different cognitive processes, serving different learning objectives, or addressing the individual representational preferences of different learners. Second, they can constrain interpretation and guide learners’ reasoning about a domain. Third, visual and verbal representations together might be suited to foster a deeper understanding than could be achieved by using just one representational format. Thus, whenever any of these functional roles can contribute to learning, representing redundant information visually as well as verbally may be advised according to Ainsworth’s taxonomy. On the other hand, if one of the representations does not contribute to learning on a functional level, it should be deleted from the instruction.

Hence, when designing multimedia materials special attention has to be paid to how to distribute information across text and graphic representations by maximizing the representations’ strengths and reducing their potential weaknesses for delivering specific information aspects. This pertains to the notion of Palmer (1978) that the represented world can be depicted in many different ways and the decision about which constitutes the most appropriate representation depends on the operations that need to be performed with it and the questions that should be answered with it.

The Role of Individual Learner Characteristics in Learning with Visualizations

In order to be effective for learning, visualizations need to be designed in a way that they can be “readily and accurately perceived and comprehended” (Tversky et al., 2002, p. 258). Moreover, the effectiveness of visualizations is affected by what the learner brings to the learning task as another crucial component of the instructional context. Visualization design needs to recognize that there is no homogenous response to graphic or textual representations and how they are, or are not, used in learning. Several learner characteristics have been suggested that may affect learning with visualizations. We will follow a suggestion by Mayer and Massa (2003), who tried to disentangle cognitive abilities, cognitive style, and learning preference along the visualizer-verbalizer dimension. These dimensions account for the fact that “some people are better at processing words and...
some people are better at processing pictures” (Mayer & Massa, 2003, p. 833).

Cognitive Abilities

Cognitive abilities related to the processing of visualizations comprise a wide range of constructs that differ not only with regard to their specificity, but also with regard to their empirical foundations. Some constructs like pictorial competence or visual literacy are rather general, whereas the impact of visuo-spatial abilities clearly depends on the type of learning tasks students are confronted with. For the former constructs, to our knowledge, no straightforward measures exist to assess them other than by observing a person performing a task that either requires or does not require this specific ability. On the other hand, there have been many different attempts to assess visuo-spatial abilities, though their literature is still equivocal (see Carroll, 1993; Hegarty & Waller, 2005 for overviews). Thus, at the current moment it is unclear how the concepts are related to each other and how they contribute to learning with visualizations independently as well as in interaction with each other. For these reasons, we will refrain from providing a comprehensive overview and instead sketch some of concepts that are relevant in the current context. The general notion is that learners who lack the abilities to process visualizations will benefit from them to a lesser extent, demonstrated, for example, in a study of science textbook instruction by Hannus and Hyönen (1999).

Pictorial competence. DeLoache, Pierroutsakos, and Uttal (2003) discuss the ability to use pictures from a developmental psychology perspective. According to their view, pictorial competence encompasses “the many factors that are involved in perceiving, interpreting, understanding and using pictures, ranging from the straightforward perception and recognition of simple pictures to the most sophisticated understanding of the conventions and techniques of highly complex ones” (DeLoache et al., 2003, p. 115). While basic elements of pictorial competence exist already in newborns, more complex skills are characterized by strong developmental shifts. In particular, the ability to understand the relationship between the representation and the referent; that is, to become symbol-minded, is clearly age-dependent (DeLoache, 1995). DeLoache’s research suggests that pictorial competence develops during the daily interactions with pictures and their real-world referents. However, it is unclear whether this kind of familiarity is sufficient to use pictures in academic contexts. According to Pozzer and Roth (2003), “most students are familiar with photographs in general; however, appropriate instructions for how to read and analyze photographs currently are not provided to them.” (p. 1092). While the concept of pictorial competency mostly refers to the interpretation of realistic visualizations, the term visual literacy denotes a more general ability of dealing with visual media of all types (e.g., video, graphics, diagrams, animation, etc.).

Visual literacy. According to Näth (2003) this concept refers to the “the ability to decode the pictorial repertoire of the media without indexical or iconic support” (p. 186). Messaris (1994) calls it “the familiarity with visual conventions that a person acquires through cumulative exposure to visual media” (p. 3). The “visual literacy” model defined by Messaris (1994) specifies different levels of visual communication that range from simple understanding to aesthetic appreciation of visual media. Despite its popularity in discussions on computer-based instruction, there is no objective assessment tool that would allow measuring the visual literacy of students. While in reading research, illiteracy can be conceptualized as a the lack of knowledge on the syntax, semantics, and pragmatics of language, there is nothing comparable in non-notational representational systems, where the interpretation of visualizations is often subjective and context-dependent. Because of this, it has also been suggested to refrain from using
the term literacy and replace it with graphicacy instead (Roth, Pozzer-Ardenghi, & Han, 2005) to get rid of the strong association between literacy and reading. The term literacy may however be adequate for logical pictures, where it can be defined as the knowledge of notational rules, allowing mapping visual features of a representation to an interpretation of the depicted part (Pinker, 1990; Shah & Hoeffner, 2002).

**Domain-specific prior knowledge.** There is increasing evidence that students’ level of domain-specific prior knowledge moderates learning from visualizations. For instance, learners with a high level of prior knowledge are better able to direct their visual attention towards relevant information (Lowe, 2003) and are less affected by a high visual complexity of the display (Lee, Plass, & Homer, 2006). Both findings support the notion that extracting information from a visualization is both a bottom-up as well as a top-down process. In the latter case, existing mental representation guide the lower-level perceptual processes and the interpretation of information acquired through them.

**Visuo-spatial abilities.** According to Carroll (1993), visuo-spatial ability is not a unitary construct; rather, it comprises five different dimensions that all make up abilities in the perception of visual input: (1) Spatial visualization is “the ability to mentally rotate or fold objects in two or three dimensions and to imagine the changing of configurations of objects that would result from such manipulations” (Mayer & Sims, 1994, p. 392) without referring to one’s self, (2) spatial orientation is the ability to imagine an object’s appearance from different view points as the observer’s body orientation changes, (3) closure speed is the ability to access representations quickly from long-term memory, (4) flexibility of closure, and (5) perceptual speed is involved in the processing of simple visual displays (e.g., quick scanning). Prior findings from other researchers have likewise established evidence for the first two dimensions (spatial visualization and orientation), while they failed to find consistent evidence for the latter three (Hegarty & Waller, 2005).

Research on spatial visualization suggests that “high- and low-spatial abilities individuals differ in the quality of the spatial representations that they construct and their ability to maintain its quality after transforming the representations in different ways” (Hegarty & Waller, 2005, p. 141). Accordingly, spatial visualization differences have been successfully conceptualized against the background of differences in working memory resources (Shah & Miyake, 1996). The role of visuo-spatial abilities has mostly been investigated in mental animation, where learners have to infer the motion of a mechanical system from a static picture. Here it has been demonstrated that learners with high abilities perform better in this task than low-ability students (e.g., Hegarty & Sims, 1994). It seems very plausible to assume that visuo-spatial abilities will show a strong influence in other tasks involving learning with visualizations. For instance, Plass, Chun, Mayer, and Leutner (2003) demonstrated that either students with low verbal or spatial abilities showed worse performance in multimedia learning than their higher-ability counterparts when receiving visual annotations, whereas abilities played no role when students were given verbal annotations.

**Cognitive Styles: Visualizers versus Verbalizers**

“A cognitive style is a psychological dimension that represents consistencies in how an individual acquires and processes information” (Kozhevnikov, Kosslyn, & Shepard, 2005, p. 710). The visualizer-verbalizer dimension (Richardson, 1977) characterizes students as verbalizers, if they rely on verbal-analytical strategies when performing a task, whereas visualizers use imagery as a predominant strategy of task accomplishment. Kozhevnikov, Hegarty and Mayer (2002) revised the visualizer-verbalizer dimension and suggest a
more fine-grained distinction between spatial and iconic visualizers (object visualizers according to Kozhevnikov et al., 2005). In a problem-solving task, they collected evidence for these two types of visualizers. Spatial visualizers, in a schematic interpretation, focused on the location of objects and on spatial relations between objects. This group also used “imagery to represent and transform spatial relations” (Kozhevnikov et al., 2005, p. 722), where images were processed analytically to infer their spatial interrelations. Iconic visualizers, on the other hand, in a pictorial interpretation, focused on vivid visual details like shape, size, colour and brightness, thereby processing these objects as a single perceptual unit. Interestingly, this distinction between the two groups fits nicely with research that focuses on dissociating the ways in which visual information is processed. This work shows that there are brain areas that either focus on processing shape and color information to determine an object’s identity (what-system) or on processing spatial and dynamic input (where-system). Moreover, current theories of visuo-spatial working memory (Logie, 1995) make a similar distinction by relating the function of the inner scribe – as one component of visuo-spatial working memory – to the processing of spatial and movement information, whereas the other part is responsible for the processing of color and shape information (i.e., the visual cache). At this moment, however, it is highly speculative if the cognitive styles identified by Kozhevnikov et al. (2002; 2005) could be linked to a different use of these brain areas or working memory systems.

Preferences for Visual versus Verbal Information

Leutner and Plass (1998) developed a method to assess preferences for verbal or visual materials by analyzing the students’ information-selection behavior. The VV-BOS (Visualizer/Verbalizer Choice Behavior Observation Scale) showed very promising psychometric properties as well as a superior validity with regard to the differential prediction of learning outcomes. Plass, Chun, Mayer, and Leutner (1998) demonstrated that students with visual preferences as assessed by the VV-BOS benefited predominantly from visual annotations, whereas students with verbal preferences gained most from verbal annotations.

Taken together, there is evidence that suggests that when investigating learning with visualizations, the students’ individual differences in terms of cognitive abilities, cognitive styles, and learning preferences need to be considered. With the current state of research, it is impossible to tell their relative influence and direction, but first studies show promising results in this respect.

CONCLUSION

When developing multimedia materials, instructional designers are faced with a multitude of decisions as to which contents should be part of the instruction, which representational format to use for these contents, and how to design these contents. For improving multimedia instructions, two conclusions can be drawn from the current paper.

First, one needs to clearly understand the learning task and relevant individual differences of the learner. From this, the relative strengths and weaknesses of different representational formats can be considered and information visually encoded in accordance with the results of this analysis. Visualizations should be used whenever their instructional functions are assumed to add considerably to the effectiveness of the multimedia instruction.

Second, the use of visualizations often seems to suffer from the resemblance fallacy (Scaife & Rogers, 1996). While it is seductive to assume that mental imagery and distal visualizations are homologous and that processing these visualizations only involve low level perceptual
processing inate to all humans, this is a grave error in instructional strategy. While perceptual processes are naturally involved in learning from visualizations, only seldom are the visualizations supposed to be taken literally. Rather, they have to be taken as representations standing for something else than what is being depicted. Identifying what is being represented based on the structural correspondences (rather than physical similarity) may cause severe difficulties for learners when studying visualizations. Thus, it should not be taken for granted that learners will extract the information from a visualization that was intended by an instructor. Rather, students need to be supported in extracting the relevant information from the visualization and guided as to how to best deploy their limited perceptual and cognitive resources. This support can be provided either by guiding learners’ attention towards its relevant aspects (e.g., highlighting) or by improving students’ competencies in dealing with visualizations. The latter comprises teaching to students existing conventions underlying these (logical) visualizations (e.g., how to read a graph, Pinker, 1990) or training them in developing graphicacy (Roth et al., 2005).

FUTURE DIRECTIONS

The final section of this paper is devoted to future directions in research on learning from visualizations. In particular, we wish to emphasize that more in-depth analyses are needed that shed light on the process of perceiving and interpreting instructional visualizations and on the visualization’s impact on performance in a variety of tasks. We believe that respective analyses would benefit from an interdisciplinary perspective, where insights from different areas (e.g., cognitive science, education, semiotics, visual perception, human-computer-interaction, human factors) could be united.

Integrating models of visual perception and cognition. Most researchers argue for visualization’s instructional effectiveness by referring to its relative ability to support higher-level cognitive processes. Accordingly, pictorial representations may support a dual coding of the information (Paivio, 1991), the construction of a mental model (e.g., Mayer, 2005), or cognitive offloading (Scaife & Rogers, 1996). These approaches pay considerably less attention to perceptual processes, thereby ignoring that these processes need to take place before higher-level cognitive processes can act upon the information that has been attended to. As a consequence, the vast literature on visual perception is largely neglected in the educational literature, even though it may provide important and novel insights into the design of effective visualizations (e.g., MacEachren, 1995). According to Anglin et al. (2004), “theory-based studies that are informed by both memory research and theories of picture perception are lacking” (p. 876).

Taxonomy research. Currently, visualizations are mostly treated as a unitary construct without taking into account the differences among them. These differences may pertain to their appearance, content, or instructional functions. Future research should attempt to develop taxonomies that will allow classifying visualizations according to these different dimensions. There have been morphological approaches that classify visualizations according to their appearance (e.g., Lohse, Biolsi, Walker, & Rueter, 1994; Twyman, 1985) however, we do not think that these taxonomies will help explaining much of the variance with regard to the instructional effectiveness of visualizations. Rather, content-oriented classifications (Bieger & Glock, 1984) or more systematic approaches to instructional functions like the one by Levin et al. (1987) seem to be more promising in this respect. The availability of such taxonomies would allow greater generalization of findings across studies than what is currently possible.

Analyzing processes and learning outcomes in parallel. While there are some theoretical
assumptions on how pictorial representations are processed, most of the existing studies have refrained from empirically investigating these processes. Knowledge on how students use pictorial representations is, however, not only necessary for theory development, it is also relevant for practical reasons. For instance, for logical pictures it is often not clear whether these representations sometimes fail to improve learning because they are designed in a bad way or because students did not understand the underlying conventions. Process-oriented data may allow identifying possible misconceptions students have regarding the meaning of the visualizations and assist in developing effective teaching strategies for using them. Despite these promises of analyzing the processes of learning from visualizations, most of the studies until now have only looked at learning outcomes in isolation. Some noteworthy exceptions to this situation come from research that has applied eye tracking methodologies (e.g., Cook, Carter, & Wiebe, 2008; Lowe, 1999). This methodology provides information on the temporal and spatial resolution of visual attention by tracking the location of the eye as a person watches a visual display. It thus offers a good starting point to analyze perceptual and possibly cognitive processes when learning with visualizations. In particular, its combination with think aloud data has been shown to be very informative (van Gog, Paas, & van Merriënboer, 2005). Hence, we suggest combining the predominant outcome-oriented research strategy with a more process-oriented strategy (cf. Peeck, 1987).

Assessment of learning outcomes. Despite the fact that core information is often conveyed by means of visualizations in multimedia instruction, most test items are presented in words only. Moreover, in class students are typically required to provide their answers verbally. Thus, most ways of assessing what has been learned from a visualization potentially requires multiple recordings of this information by the learner to accomplish a task. In particular, non-sequential information needs to be transformed into a sequential format to provide a verbal answer (speaker’s linearization problem; Levelt, 1981). Moreover, according to the verbal overshadowing effect (Melcher & Schooler, 1996) it may well be that asking learners to provide verbal answers when assessing their learning outcomes may interfere with the “visual” knowledge stored in memory and make this information less accessible. Hence, verbal tests may inadequately reflect students’ acquired knowledge. Moreover, interactions between the instructional format and the learning objective should be considered more thoroughly in future research. A larger variety of test formats might be more apt to account for the complexities underlying learning with visualizations. Accordingly, Joseph and Dwyer (1984) found positive effects for illustrated text compared to text-only versions only in a drawing test, but not in a comprehension test, suggesting that the type of test moderates the effectiveness of instruction. Furthermore, Levie and Lentz (1982) showed in their review on learning with visualizations that visualizations improved recall more in delayed tests than in immediate tests. It might be that the advantages of visualizations relate to dual coding in memory and creating more associations to other long term memory content—a pay off only seen in situations that impose higher demands on memory and understanding.

REFERENCES


A. Houghton (Eds.), *The psychology of illustration* (Vol. 2, pp. 159-188). New York: Springer.


Theoretical and Instructional Aspects of Learning with Visualizations


Kozhevnikov, M., Kosslyn, S., & Shephard, J. (2005). Spatial versus object visualizers: A new characterization of visual cognitive style. Memory & Cognition, 33, 710-726.


ADDITIONAL READINGS


