EFFECTS OF PICTURES ON MEMORY & LEARNING

Walter KLINGER
ウォルター クリンガー

Foreign language instruction material has become increasingly highly visual--from pictures on flashcards for vocabulary learning, to richly illustrated textbooks, to multimedia software, to films and movies. In this article, we consider what the advantages and disadvantages may be of using visual material. What happens when we look at pictures, or at written words, or at pictures and words together? Is there a further effect when we hear words spoken along with visual or written material? We find evidence that visual material has strong influences on memory and learning, but its effectiveness for language learning depends on the goal of the instruction as well as student learning abilities and preferences.

The goal of my classes for first-year college students is to encourage and develop ability in speaking and hearing English, as contrasted with ability in reading and writing, so I often use visual material to divert students’ attention away from the written word. I sometimes use movies and songs with video clips and I have developed a number of games which were described in previous articles (Klinger 1999, 1998) and are available online at http://www2.ice.usp.ac.jp/wklinger/QA/cardgameshome.htm/.

One game I use is a set of several dozen drawings, without captions, of everyday events and activities. Players in turn put down a card to make a continuing story of what happened one day. The play is noisy and lively since a player with a card that is more related to the exposed card can challenge someone who tries to put down a less related card. Almost always, however, the sentences the students spontaneously say are grammatically incorrect, so in some classes I ask the students to write out sentences beforehand and check them for correct grammar and usage.

Another game is a set of question and answer cards. Each pair of cards has the same picture but different captions; the goal of the game is to find who in your group of players
has the answer card for your question card. This game can also be played without written captions, so that players supply their own sentences, but for a study to find out to what extent pictures might make the words and sentences memorable, I used a version with prepared captions.

One day, I showed the cards on an OHP and the students repeated the sentences after me. While playing the game, the student with a question card might need to repeat the sentence between 1 to 4 times before finding who had the answer card. The game was played at least 3 times, followed by games such as “Slap!” or “Concentration” where the sentences were said a few more times. In the next class the following week, I handed out sheets showing just the pictures and asked the groups to write whatever they remembered, even if it was only one word, or even to write in Japanese if they couldn’t remember the English.

From this experiment, I concluded that the pictures themselves were quite memorable, but if the hope was that pictures would help sentences and spelling to be memorable, they were not especially effective. A few of the results are as follows:

One set of cards read: Q: What are you going to give your mom for Mother’s Day? A: I haven’t decided yet. Maybe flowers!! Out of 50 test papers, 34 had complete, if not necessarily grammatically correct, sentences. “Give” appeared 16 times. “Present” appeared in 24 responses, sometimes as a verb, as in “What did you present Mother’s Day?” or as a noun in phrases like “decided a present.”

Q: What are you going to do this weekend if it rains? “If it rains” was given in 10 responses, and “if it will rainy day,” “if it is rain,” “if it is rainy,” “if it rain,” and “if it’s rain” were given once each.

This question asked: Have you ever gone all the way up Mt. Fuji? “Climbed,” including “crimbed” and “clumbed,” was written in 16 responses, “seen” in 6 responses, “been to” in 5 responses, but “gone up” in only one.

Q: Do you know what apple juice is made from? A: Apple juice is made from apples, of course, silly! A text box gave the Japanese translation for “silly”--馬鹿chan! 37 responses were full sentences. “Silly” was given 18 times, along with variations like “shilly,” “stupid,” “foolish,” “foo,” and “Apples!” “Made from” was given 33X; “made of” 4X, including “It make from apple,” “It made from an apple,” “What is made from apple juice?” “What apple juice is made from?”
and “What does apple juice make from?”

The question for this card read: Do you like bamboo shoots? In Japanese, they’re called takenoko. The answer card read: Sure! They’re a delicious treat in the springtime! Have you ever gone and dug them up? The English word “bamboo” was probably known to most students, but “shoots” was probably new vocabulary. “Dug” was written in a distinctive font, as were many other words on other cards, which I hoped would act similar to a picture effect and make the words memorable. During the OHP presentation, I also drew attention to the word, pointing out that “dug” comes from “dig: dig-dug-dug.” However, only 5 responses mentioned “dug.” “Bamboo” was spelled correctly 19 times, “banboo” was written 12X, and other spellings appeared 13X, like “bumboo,” “bumboo,” “bamebo,” and “banbboots.” “Shoots” or “shoot” appeared 18X, besides forms like “soots,” “shuits,” “shut,” and “shorts.” “Shoot bamboo” was written once. “Takenoko” was given 15X. 28/50 papers had full sentences.

If I had told the students from the outset that they would be tested, the results could well have been better, as the expectation of a test can provide motivation to remember (Hill 1966:47). I also did not strongly insist that the “test” paper be filled out, so often pictures were ignored or given minimal responses. I let students work in groups of 2-6, so the results are collective rather than individual. While the students were writing out their responses, I asked several groups if they remembered seeing a particular picture, to which they replied that they did but that they couldn’t remember the caption. The pictures often enough did prompt the gist of the sentence, if not the exact grammatically correct sentence, so to that extent the pictures had some effect on memory. Mehler (1963:346) said the message of the sentence is the meaning, not the details of the grammar of the sentence, and people can rephrase better than they can remember grammar details.

Considering that the students have fairly low ability in grammar, I did not expect pictures to suddenly produce grammatically accurate sentences. The spelling mistakes were a bit disappointing, however, but the pictures might actually have been a distraction from students noticing the given correct spelling, as we will consider below. The pictures often elicited comments on their charm and attractiveness, so they also function as motivators by providing interest, though that does not mean that there is necessarily any overflow of attention from picture to text.

I find that pictures are very effective on their own, without any accompanying text, in
providing motivation to the students to try to make their own self-generated sentences. Some groups of students enthusiastically started discussing and writing their own captions during the “testing” class, and were disappointed when I told them that at this time I just wanted to find out what they could remember of the given captions. Some students also have told me that this game using prepared captions was not as enjoyable as the games where they could make their own sentences. As material that promotes spontaneous speaking, even with sentences that are ungrammatical -- but hopefully will improve with repeated attempts, pictures are very useful teaching tools. I wrote on the theme of the importance of spontaneous speaking for language learning in a previous article (Klinger 1999).

To further understand what happens when pictures are used as learning material, we review reports from the literature. After defining the different types of illustrations, we consider theories of how the brain processes pictures and words, and look at studies of how pictures can help --or hinder-- learning.

Types of Visual Representations

Physical objects and events that can ordinarily be seen and actions that can be performed in the everyday world can be represented artistically, in drawings, sketches, caricatures and paintings, in still photography and in motion pictures, or in computer simulations. These illustrations are what I refer to in this article as “pictures.” Objects and events may also take the form of mental images in the mind, i.e., “imagery.” In many cases, objects and actions may also be mimed in gestures.

Contrasted to these real-life representations are illustrations that visually describe hierarchies, sequences of events, or relationships between concepts or statistics. They take forms such as graphs, diagrams, geometric models, line charts, bar charts, scatter plots, maps, network charts, flow charts, time lines, data tables, trees, webs, matrixes, pie charts, Venn diagrams, classifications of plants and animals, sequences of events such as chemical reactions and nutrient cycles, diagrams of sentence structure, and other such “graphic organizers” or “visual analogies.” Levie (1987) separates visual information into further categories. Butler (1993) also offers a more detailed classification and surveys the increased use of visual information in psychology textbooks and journals over 50 years.

Written text is at a far end of the continuum of object-picture-word, insofar as a word is abstract compared to the tactile object it names (Giddan & Giddan 1984:18). Yet words
are also partly visual information. Text can be made more visual when it is arranged in a non-linear way on the page, or set off in a distinctive font. Even text that is presented in standard paragraphs may be highlighted in color, circled or underlined or italicized as key words or alert words, or rewritten as an outline or summary.

**Different Processing of Words & Pictures**

But words, when a literate person sees them, are not only visual but also auditory. Most linguists believe that representations of words in the mental lexicon are coded phonologically, though some say that there is also a “separate visual-orthographic lexicon” (Frost 1988:741). That is, when we see a word, we probably look it up in a talking mental dictionary based on what we know of what sounds the letters make, or perhaps we look up the word in a visual but silent mental dictionary based on what we know of what the shapes of the letters mean.

Frost’s (1988) study suggests that printed words are immediately and automatically recoded into internal phonetic form. McKoon (1977: 247) describes research by Kintsch and Craik & Lockhart that says that text is encoded and analyzed in several levels of processing: graphemic, phonemic, lexical, syntactic, and semantic, with each level leaving traces in memory. Bajo (1988:581) says that a picture must be meaningfully identified (access its semantic representation) before it can be named (access its phonetic or name representation), while a word may be named before its meaning is activated (it does not necessarily access its semantic representation when read). Pictures access meaning directly and phonemic features indirectly. Words access phonemic features directly, and meaning only after some phonetic processing.

That there are different processes apparently at work when we look at pictures and words has provoked a debate of whether the brain stores information about pictures and words separately. The “Dual Coding Theory,” as proposed by Paivio, says that there are two separate but interconnected knowledge systems: a verbal system, specialized in dealing with sequential information in discrete linguistic units and abstract concepts, and a visual imagery system which encodes, stores, organizes, transforms and retrieves spatial information about concrete objects and events (Hodes, 1992:47). Marschark & Paivio (1977:228), referring to mental rotation studies, believe that there are “perceptual-analogue representations in long-term memory, which encode various attributes of the objects themselves.” In the Dual
Coding approach, the verbal subsystem contributes to logic, order, direction, and organization of thought. The nonverbal subsystem consists of “holistic nested sets of information,” is “relatively free of logical constraints,” and is better at coping with concreteness, spatial information, and the analysis of scenes (Gambrell & Brooks Jawitz 1993:266).

Levie (1989:10) says even dual-code models might be too narrow, considering evidence that human faces might constitute a special class of stimuli which are handled by yet a third memory system. Rolls (1995:177) describes the specialization of function in the temporal lobe visual cortical areas: Some areas are multimodal, responding to visual and/or auditory inputs, some are unimodal visual, some are specialized in analysis of moving visual stimuli. A large number of neurons are devoted to processing information about faces, reflecting the importance of face recognition and expression identification in primate behavior.

Rejecting the theory of Dual Coding are Rumelhart, Schank, Anderson & Bower, Pylyshyn, Nelson, and others, who support a “unitary conceptual code” (Potter, Valian, & Faulconer 1977:2). Bartlett, Till, & Levy (1980:447) concede that verbal and nonverbal information is to some degree separable in memory, in that people can control the extent to which they use one or the other in recognition, but it does not necessarily follow that verbal and nonverbal information must be organized in two different memory traces.

This debate about how pictures and words are similarly or differently processed in the brain is relevant to the topic of whether or not to use pictures in teaching material, and how to use them. Molitor, Ballstaedt, & Mandl (1989:9) observe that if dual coding is correct, and different processes are at work, then different teaching material should be used, whereas if there is a single code, then the presentation mode is not important.

A number of widely acknowledged phenomena suggest that pictures and words are processed differently. Research shows that most individuals read with the left-brain hemisphere, but the right hemisphere can understand some written information referring to “concrete, picturable objects” (Benson 1981:83). Damage to the left hemisphere resulted in Japanese subjects having problems understanding syllabic kana, which have more auditory aspects, while damage to the right hemisphere resulted in problems understanding kanji, which are more nonphonetic and visual. (Benson 1981:82).
Different Processing of Words & Pictures

One often observed effect suggests that pictures and words do not share a common processing system: Pictures are inevitably remembered better than words on tasks of recall and recognition (Park & Gabrieli 1995:1593). This “picture superiority effect” is an “established memory phenomenon,” in that experiments have repeatedly shown that “memory for pictorial stimuli is extremely accurate, durable, and extensive compared to that for verbal stimuli” (Noldy, Stelmack, & Campbell, 1990:417). In Shepard’s (1967) experiment, subjects recognized old word stimuli 90% of the time, sentences 88% of the time, and pictures 98% of the time. Pictures become even easier to remember when the objects are not just side-by-side but are shown interacting, e.g., a car crashing into a tree (Wollen & Lowry 1971:283).

A theory to explain why pictures are memorable says that the processing of pictures in the brain needs “additional allocation of attentional resources or effort” (Noldy, Stelmack, & Campbell 1990:418). Noldy, Stelmack, & Campbell’s (1990) EEG recordings of brain ERP (Event-Related Potential) waves showed that it took longer to name a picture than to read the verbal label of the picture. Park & Gabrieli’s (1995:1589) participants also named pictures more slowly than they read words. Investigations of elementary learning processes, such as free-association reactions to words, drawings, and objects, have since the 1940s found a longer reaction time to pictures than to words (Otto 1962).

Pictures are more complex than the words that label the pictures, so more time and attention is needed to identify, or “name,” a picture. We spend more time looking at pictures (or real-life objects) before we can name them, so we remember pictures better. We spend less time looking at words in sentences, so we don’t remember the sentences exactly—though we remember the gist. Pictures are also more distinctive and more unique than the words that label them, which further makes pictures more memorable.

Adults gain experience with reading words, so reading becomes faster and faster. Processing for words becomes automatic through practice; it is “performed without the necessity or benefit of intention” and “requires minimal allocation of attentional resources.” “Lexical and phonological access may be automatic for words but not for pictures,” continue Noldy, Stelmack, & Campbell (1990:418). When their subjects were looking at words or pictures, there were differences in ERP waves, which reflects the differences in the cognitive processing of pictures and words. When their subjects were told to pay attention to words or
pictures, the ERPs for pictures changed, but not the ERPs for words. This result suggests that the early processing of words is automatic, i.e. unaffected by attention, while the early processing of pictures is controlled, i.e. influenced by attention (Noldy, Stelmack, & Campbell, 1990:426).

“The more frequently a word is read, the more efficient its mapping between the orthographic input lexicon and the semantic system,” say Borowsky & Besner (1993:833). Words can be read faster than pictures can be named, Seifert (1997:1106) confirms, because words “activate their lexical entries automatically. Then, only after that do they activate their meanings in semantic memory. . . . Pictures activate information in semantic memory first, and then they activate lexical information.” That is, when we look at a picture, we see what it “means” before we name it or identify it using a word.

Gibson’s (1966) theory of cognitive visual perception says that the information we perceive in the environment is so rich that we do not even have to process it--the environment explains itself and almost “tells” us what it is (Doerr 1999). When we see a chair, we understand its meaning without needing to name it. We do, nonetheless, have to learn the meaning of the chair, its function or physical composition. As Ogasawara (1998) points out, “no one can understand the pictures(s) and/or figure(s) without any preconcept or prelearning.” While naming pictures is slower than reading words, pictures can be understood quickly when we don’t need to verbalize what we are looking at.

Seifert (1997:1115) found an advantage for pictures when subjects had to identify pictures as fitting into categories, compared to identifying words in categories (e.g., engine, car; knife, spoon; tail, dog; toaster, bread; chicken, eggs). Identifying pictures as belonging to a category was 21-57 milliseconds faster than identifying words; a small difference but sufficient to demonstrate that the phenomenon exists. However, pictures were not faster than words if subjects had to decide if the pictures or words did not belong in a category (e.g., toaster, cow; eggs, lamp).

**Learning By Making Connections**

We learn from experience what things belong together in categories; we have learned the connections between different concepts. In picture and word experiments, this effect of being able to quickly spot the relationships between pieces of information is known as the priming or context effect, or the “expectancy mechanism” (Borowsky & Besner 1993:813).
Words as well as pictures are read or understood faster when preceded by a related context. For instance, it is easier to identify the word “butter” when it comes after “bread” than when it comes after “doctor” (Borowsky & Besner 1993:813).

Long-term memory, says one theory, stores information in “schema,” which is a “cognitive construct that permits one to treat multiple elements of information as a single element categorized according to the manner in which it will be used” (Marcus, Cooper, & Sweller 1996:49). For example, we can develop a schema for trees, and recognize a single instance of a tree as belonging to the category of “tree,” despite each tree being different in many ways. Reading involves automation of schema for letters and then words, say Marcus, Cooper, & Sweller (1996:50). Knowledge is remembered in webs of schemata; the more information is interconnected, the more it is readily recalled (Saunders, Wise, & Golden 1995:42). Poor learners, says Patton (1991:18) are usually poor organizers who have little ability to devise conceptual patterns.

A “multimodal theory of learning,” like the Dual Coding theory, says learning is more effective when the learner uses more than one sense modality, for instance, verbal and visual processing, and when connections are clearly made between the information contained in each modality (Mayer & Sims 1994:390). The “theory of generative learning” proposed by Wittrock says that reading comprehension develops when readers build relationships between the text and their own knowledge and experience: “The meaning of a text is not inherent in the print on the page but is the result of constructive processes that the reader brings to bear on the message” (Gambrell & Brooks Jawitz 1993:266).

Presenting Words & Pictures Together

This way of learning, by making connections between pieces of information and understanding their relationship, suggests that presenting pictures and words together so as to create connections between them will similarly be beneficial. A large body of research shows that learning can be affected positively when text and illustrations are presented together. Moeser & Bregman (1973:91), for example, report that subjects learning a miniature artificial language did not manage to learn any syntax if 3,200 sentences were presented alone, but showed excellent results if sentences were accompanied with pictures.

Verbal-visual associations create personal associations for unfamiliar words, say Hopkins & Bean (1998:275). Many experiments conclude that if pictures and text are
presented together, they should be presented simultaneously rather than separately, so that the two representations are in memory at the same time, and “construction of referential connections” can be done immediately (Mayer & Sims, 1994:391).

The ease in which information is understood is influenced by “the intrinsic complexity” of the information and “the manner in which information is presented” (Marcus, Cooper, & Sweller 1996:50). Text is processed sequentially, which is “ill-suited for tasks where multiple constraints must be considered simultaneously,” so diagrams gain effectiveness “by reducing cognitive load by making spatial relations explicit” (Marcus, Cooper, & Sweller 1996:52). When studying the anatomy of ear, for example, an illustration is superior to text because the topic involves concrete spatial relationships (Iding 1997:22). In a study of very complex information, Chanlin (1997) found that an animation about recombinant DNA technology produced good results, but it required much mental effort, with students often or constantly referring to written explanations while viewing the animation.

When text information is simple, however, diagrams will not make learning easier. The advantage of diagrams over text “should be lessened when information can be processed serially because element interactivity is low” (Marcus, Cooper, & Sweller 1996:60). This is borne out in studies like Kliese & Over’s (1993) tests of students’ understanding of the operating principles of pumps, thermostats, and electrical relays: students showed no greater knowledge when text was accompanied with illustrations of parts and steps, compared to text alone. Adding pictures and graphics can increase learning speed and help learning and retention if they supplement the text in some meaningful way, but print information is necessary for accuracy, say Ellis, Whitehill, & Irick (1996:130).

Some research has been done on how the style of the text in captions accompanying pictures and graphs affects memorability. A “selective-attention theory,” says Iding (1997:2), suggests that readers pay more attention to information presented in question form, so the chance of recall should be increased. Some experiments show that questions in science charts increase memorability, but Iding’s (1997) own experiment failed to show that students perform better when captions were questions about the illustration rather than descriptive information. Bernard (1990:224) found better results when the caption was a short instructive direction of what to do while looking at the picture, e.g., tracing the route of the pain process in the nervous system, compared to a descriptive caption.
Even more effective than memory of pictures is memory of interaction with operating and assembling an actual device. In their study of structural knowledge of how to build a crane, Ellis, Whitehill, & Irick (1996) found that getting hands-on experience was better than pictures that only provided information.

To learn a subject, students need to be active and “learn by doing.” In multimedia combinations of video, sound, text, animation, and graphics, students can manipulate information in a fairly realistic way. Students can also decide for themselves the type of information they prefer to access and the order in which to process information. Wittenmark, Haglund, & Johansson (1998) describe the benefits of dynamic pictures compared to static pictures in their web-based, computer interactive modules where users can improve understanding by changing a parameter and immediately seeing its influence. A “cognitive theory of multimedia” says that meaningful learning requires that the learner engage in five active cognitive processes: selecting words, selecting images, organizing words, organizing images, integrating words and images (Mayer, Bove, & Tapangco 1996:65).

Student Abilities & Learning Styles

The effectiveness of pictures is further qualified by student ability and preferred learning style. Kliese & Over (1993:185) suggest that diagrams facilitate understanding only for students with low prior knowledge and even then, students with little background can be overwhelmed by complex graphics if the connections are not obvious. Kliese & Over (1993:185) further say that, as individual student learning styles and strategies differ, so the results of how effectively they interrelate text and illustration will also differ.

Ollerenshaw, Aidman, & Kidd’s (1997) study shows that illustrating and labeling parts and operating stages of a device improves recall, and results are even better for those students who have low prior knowledge. They too suggest that results may differ according to differences in student learning styles, and give a classification developed by Biggs. “Surface approach learners” implement strategies both to avoid working too hard and to avoid failing. “Achieving approach learners” aim for high grades and satisfaction from high grades. “Deep approach learners” study to satisfy curiosity and to extend their knowledge to other areas. In their test of understanding of how pumps work, Ollerenshaw, Aidman, & Kidd (1997) found that surface learners recalled less, but they probably benefited from visuals more than other kinds of learners, perhaps because they needed to “exert less effort”
by looking at pictures compared to studying text. Text plus animation made little or no difference for knowledgeable or deep learners.

In their study of a German language-learning class, Plass, Chun & Leutner (1998) showed how vocabulary learning was affected by individual preference of learning style, classified as “visualizers” or “verbalizers.” Using a multimedia computer story, college students could look up marked words by hearing the pronunciation plus seeing either a written translation of the word, or a picture (half the time a video clip) depicting the word. Visualizers recalled illustrated words much better than unillustrated words. Verbalizers recalled illustrated propositions only slightly better than unillustrated ones. The best results occurred when both illustrations and translations were looked up, with only a small difference in results for visualizers and verbalizers.

Di Vesta, Ingersoll, & Sunshine (1971) administered a battery of tests on college students to determine their differing verbal and imagery skills. Verbal or symbolic imagery tests included Reading Comprehension, the Scholastic Aptitude Test (verbal and math scores), a Vocabulary Test (verbal analogies, definitions, opposites), and a Remote Associates Test (the ability to make links in groups of words). Visual tests included Space Relations (imagining rotating pictures or patterns), Space Thinking, and Gottschaldt Figures (tracing figures).

Visualizers and surface learners both benefit from pictures because their brains need to “exert less effort.” Verbalizers benefit from text because their brains need to exert less effort. Reichle, Carpenter & Adam (2000) found that visualizers and verbalizers function better when they can choose their preferred style experiment, in an experiment using functional Magnetic Resonance Imaging to “examine the relation between individual differences in cognitive skill and the amount of cortical activation engendered by two strategies (linguistic vs. visual-spatial) in a sentence-picture verification task.”

The verbal strategy produced more activation in language-related cortical regions (e.g., Broca’s area), whereas the visual-spatial strategy produced more activation in regions that have been implicated in visual-spatial reasoning (e.g., parietal cortex). These relations were also modulated by individual differences in cognitive skill: Individuals with better verbal skills (as measured by the reading span test) had less activation in Broca’s area when they used the verbal strategy. Similarly, individuals with better visual-spatial skills (as measured by the Vandenberg mental rotation test) had less activation in the left parietal cortex when they used the visual-spatial strategy. These results indicate that language and visual-spatial processing are supported by partially separable networks of cortical regions and suggests one basis for strategy selection: the minimization of cognitive workload (Reichle, Carpenter &
This preference in using visual information rather than written information may be innate or it may be a choice, but it also has cultural determinants. Historically, pictures were a medium for illiterate people (Molitor, Ballstaedt, & Mandl 1989:4), and some people may regard pictures as being inferior or childish compared to text. Goldsmith (1987:53) notes that some adults were embarrassed to be seen with illustrated reading books during a UK adult literacy campaign in the 1970s. Sarkar (1978:175) notes that in some Arab and African societies, “serious books” never have pictures since pictures are considered “kid stuff.” On the other hand, the visual has become such “a dominant and defining element in our literacy and culture,” as Woodward (1989:101) says, that buyers/readers will not open a book/magazine that does not have visual appealing material.

Di Vesta, Ingersoll, & Sunshine (1971:477) describe a “Social Desirability” test, to see how having visual talent rates compared to having verbal ability. For some people, they find, being able to experience vivid images and rich fantasies is preferable to having a high IQ. Some researchers, they say (Di Vesta, Ingersoll & Sunshine 1971:472), believe that “permissiveness in child-training” has influenced the continuing preference by some adults of imagery and language habits related to imagery.

**Advantages of Pictures for Children**

For adults, say Di Vesta, Ingersoll, & Sunshine (1971:478), use of imagery is a strategy preference, while for children it is a skill. Teaching children to construct mental images as they read enhances their abilities to “construct inferences, make predictions, and remember what has been read,” say Gambrell & Brooks Jawitz (1993:265). Their study (Gambrell & Brooks Jawitz 1993) showed that second and fifth graders who were told “Remember to make pictures in your head” outperformed the control group who were instructed to “Read to remember.” Mental imagery, they found, by encouraging an active engagement with the text, was even more effective than illustrations for reading comprehension, as the children recalled more story structure and wrote more complete stories.

Allender (1991) taught English vocabulary to 8-11 year-old children in Japan who listened to spelling with eyes closed, calling out the word, and writing the word with eyes closed. The children also imagined visual signposts with words spelled on them on the road while walking to a store. Cloze test results seemed “slightly better than other semesters and
other classes, but certainly not dramatically so” (Allender 1991:52).

Levin (1981) reviewed 204 studies that invariably demonstrated that pictures helped children understand prose better, with up to 40% more recall of information. Pictures help children identify unfamiliar words during reading, and enhance comprehension and retention of text, but when learning to recognize and remember the written representations of words, some studies suggest it may be more efficient when written words are presented by themselves rather than along with pictures. Solman & Wu’s (1995) study showed that it took children longer to learn single written words when taught with associated pictures. (Solman & Wu 1995:227). According to the “focal attention hypothesis,” pictures capture attention, and so they distract from written words. When the child sees a picture, it “is likely to trigger automatically the verbal response. Consequently, the attempt to focus children’s attention on the written words is ineffective” (Solman & Wu 1995:229). They recommend that a picture be presented only after a response to the written word has been made, as feedback “to correct a child’s incorrect response, reinforce a correct response, or ascertain the meaning of an uncertain word” (Solman & Wu, 1995: 241).

Though they found “an absence of a direct instructional value of pictures,” Solman & Wu (1995:234) refer to numerous studies that suggest that one great benefit of pictures is to enhance the incentive to learn. The affective factors such as attitude, motivation, and interest may lead learners to pay more attention and put in more effort, which may lead better learning. Pictures may also “enhance the long-term retention of the words” since the “dual-coding effect” and the “greater effort” made “to process information and establish the relationship between the various sources of information” may make learned information “more resistant to memory loss” (Solman & Wu 1995: 235).

**Advantages of Text for Adults**

Children are allowed and even obligated, because of their low or nonexistent ability with reading text, to use pictures and imagery. While some adults have a preference for visual over text information, the typical development from child to adult still seems to be that, as the child begins to think through the use of “symbols,” “ikonie” imagery gradually fades; that is, “verbalization is substituted for imagery” (Di Vesta, Ingersoll & Sunshine 1971:472). Verbal skills eventually become “free of dependence on more concrete visual images” (Moeser & Bregman 1973:92).
This developing ability of coping with written text for adults often leads to advantages of text over visual, and as text becomes both easier and more satisfying, some adults show a preference of text. In my own classes I find that some students say that, while they enjoy the picture games, they prefer the games we play that involve spelling and making and finding written words.

In Otto’s (1962) study, 20 college students were given picture and word stimuli, and asked to tell how the pictures or words were alike. Otto was trying to get responses like “soft,” “round,” “sweet,” “toy,” and “hole.” The text words produced more responses for the adults. For 4th Grade children, on the other hand, the pictures evoked more responses even among good readers. Poor readers made more no-responses to words but also to pictures.

van der Molen & van der Voort (1997) report on several studies that show that adults recalled print news better than television news, probably because print offers adults more control over information processing. Grade 4 and 6 children in their experiment, on the other hand, recalled TV news better than the identical print news, regardless of whether they expected to be tested for memory or not.

Many experiments with pictures and text use just single words, but in real life we usually read longer information. In paragraphs, compared to “isolated words in lists,” Marschark (1985:744) says, “images don’t serve the function of higher order organizers.” Marschark (1985) found no difference in recall and comprehension of high or low imagery paragraphs of six sentences, and actually found more abstract paragraphs were remembered better than more visual sentences. Johnson & Steele (1996:354) point out that, compared to the highly visual words the children learn first, difficult vocabulary often expresses relationships and other abstractions, e.g., “stipulate,” “augment,” and “unsubstantiated.” To learn words like these, we can try to create visuals, or “relate the word with visual or acoustic properties,” through techniques like mnemonics, but they are not always helpful.

Hearing, Words & Pictures

We also need to consider the effects of spoken information along with pictures and text. Children’s emergent reading, says Elster (1998:43), is influenced by the number of times they hear a book read, and then read on their own. Their control of the linguistic code begins from shared readings, where children” link the meaning of words in books to their own experiences and knowledge of the world” (Elster, 1998:45). Pictures are a “scaffold”
which foster attention and prompt talk and comments (Elster, 1998:68), but language skills proceed from hearing and speaking, in activities like naming objects to describing a picture and following verbal commands such as “Put the pink pig next to the horse” (Debaryshe 1993:457). “Procedures for meaning-making,” say Hester & Francis (1995:85), are learned through oral exchanges and social interactions in groups, and pictures are beneficial for this purpose as they “invite children as co-participants” (Hester & Francis 1995:66).

In van der Molen & van der Voort’s (1997) study, children recalled TV stories (visual+hearing) better than if they read the stories or if the stories were read out loud to them. For adults too, adding sound to pictures may make the information more memorable. Baggett’s (1979:333) test of the recall of information presented in a text and in a movie showed that the text was forgotten faster. Baggett (1979:354) suggests that film produces a “more emotional reaction” and this somehow impacts on memory.

Berkay (1995) wondered if adding written captions (presented at 150-180 wpm) to a videotape (visual + audio) of “fairly challenging information on psychology” to university students would be distracting or beneficial. The “Between-Channel Redundancy” theory, again like the Dual Coding theory, suggests that when information is redundant between two information sources, in this case, captions and dialog, comprehension should be greater than when only one source is used (Berkay 1995:288). Berkay (1995) found that adding captions does not increase comprehension, but neither does it diminish comprehension, despite some people’s feeling that captions are distracting.

One theory proposes that, as far as pictures and words are both types of visual information, it might be better to supplement visual material with audio material rather than more visual material. When both pictures and words are held in visual working memory, say Mayer & Moreno (1998:318), not enough working memory is left over to build connections between words and pictures. Mayer & Moreno’s (1998) students viewed a silent animation on the formation of lightning, with either concurrent spoken narration or concurrent on-screen text. The animation + narration group outperformed the animation + text group. Working memory may have “separate visual and auditory channels” so that if animation+text are used, visual working memory is likely to become overloaded, or, due to a split-attention effect, fewer pieces of relevant information can be selected.

Eddy & Glass (1981) found that hearing “high imagery” sentences produced better understanding than reading them. Subjects were asked to rate as true or false sentences such
as: “A tic-tac-toe game is played with four lines.” “A tic-tac-toe game is played with five lines.” “A row boat comes to a point in the back.” “A row boat comes to a point in the front.” Hearing the sentences produced faster answers than reading them. Glass, Millen, Beck, & Eddy (1985:449) confirmed that high imagery sentences were more difficult to understand when read than when listened to, and noted that all “sensory modalities” can have “imagistic representations.” When we hear or read “Milk chocolate is sweeter than dark chocolate” or “A typewriter is louder than a ticking watch,” a number of multisensory images may arise, not only visual images.

**Conclusion**

Pictures, we have seen, are often helpful in introductory material for students with low prior knowledge of a study topic. This group of students includes children learning to read and students beginning foreign language studies. But pictures can also distract from learning to read text. The critical factor in learning to read alphabetic scripts, say Huang & Hanley (1995:74), is to “isolate and manipulate individual phonemic units,” so the most important visual skill is “orthographical rather than pictorial or logographic.” However, excellent visual memory skills can be a “significant advantage” in learning to read non-alphabetic scripts like Chinese characters (Huang & Hanley 1995:93).

Memory seems to store information in webs of relationships, so the benefit of placing pictures together with text comes from the possibility of making connections between the two types of information. Studies show that foreign words are learned more easily when associated with actual objects or a corresponding picture, or imagery techniques, or translation (Plass, Chun, & Leutner 1998:26). But it is doubtful whether pictures or imagery can help much with non-visual, abstract vocabulary or with sentences and paragraphs that require and understanding of the abstract rules of grammar.

Some people have a natural or a learned preference for visual material, and learning requires less effort when they can use their preferred material. Children say TV is an “easy” medium and print a “tough” medium; which means they need to invest less mental effort in trying to process information from TV (van der Molen & van der Voort 1997:84). Interestingly, one might expect that the extra effort required to understand printed text should result in print being more memorable, similar to the extra time spent processing pictures helps make pictures more memorable. However, as Salomon said (quoted in van der
Molen & van der Voort (1997:84), learning explicit factual information doesn’t require deep processing, so results may be equal from TV or print, whereas inferential learning requires deep level processing and best results may be from print.

Pictures can be understood very quickly. The global meaning or the gist of a picture can generally be attained in an easy and rapid way, in as little as 300 ms (Peeck 1993:228). This “gives the illusion of full understanding” and so “students often deal with illustrations in a superficial and inadequate way” (Peeck 1993:228). This is a topic will be further developed in a future essay, along with the effectiveness of gestures in language learning.


