Supporting Visual and Verbal Learning Preferences in a Second-Language Multimedia Learning Environment

Jan L. Plass
University of New Mexico

Dorothy M. Chun and Richard E. Mayer
University of California, Santa Barbara

Detlev Leutner
Erfurt University of Education

English-speaking college students who were enrolled in a German course read a 762-word German language story presented by a computer program. For key words in the story, students could choose to see a translation on the screen in English (i.e., verbal annotation) or view a picture or video clip representing the word (i.e., visual annotation), or both. Students remembered word translations better when they had selected both visual and verbal annotations during learning than only 1 or no annotation; students comprehended the story better when they had the opportunity to receive their preferred mode of annotation. Results are consistent with a generative theory of multimedia learning that assumes that learners actively select relevant verbal and visual information, organize the information into coherent mental representations, and integrate these newly constructed visual and verbal representations with one another.

Current developments in information technologies have resulted in rapid advances in the application of instructional and educational technology. There seem to be, however, only small advances in corresponding basic research on the psychological principles of human learning in a multimedia learning environment (Kozma, 1991). This article deals with two of these psychological principles. On the one hand, multimedia learning requires the learner to process information presented in different modes (e.g., in a verbal and a visual mode). Thus, cognitive psychological theories on processing information such as Mayer's (1997) generative theory of multimedia learning are relevant. On the other hand, the term learning environment points to the fact that learners in such an environment are usually allowed to decide for themselves the type of information they prefer to access and the order in which they process different types or modes of information. Thus, psychological theories on individual differences in learning preferences come into play. One dimension of such theories that seems to be of specific relevance for learner-controlled information seeking and information-processing behavior in a multimedia learning environment deals with the distinction between visualizers and verbalizers (Keefe, 1989; Kirby, 1993; Kirby, Moore, & Schofield, 1988; Richardson, 1977).

Following Mayer and Sims (1994) and others (e.g., Corno & Snow, 1986; Jonassen & Grabowski, 1993) who called for more research on the role of individual differences in multimedia learning, the present article is concerned with the question of whether visualizers and verbalizers differ in their behavior in a multimedia learning environment and how the differing needs of visualizers and verbalizers can be supported to improve their overall learning outcomes. The learning environment used here is an authentic second-language learning situation and requires the learner to read a literary text written in a foreign language and annotated with information in the form of spoken words, translations, pictures, and video clips. This article is concerned with two different levels of learning: acquisition of single vocabulary items, which may be considered a form of rote learning, and comprehension of the basic events in a text, which may be considered a form of constructive learning. Of specific concern is how individual differences, particularly in the learning preferences of visualizers versus verbalizers, can be accommodated in an attempt to shed light on the question, "For whom is multimedia instruction effective?"
Extensions of a Generative Theory of Multimedia Learning to Second-Language Learning

This study builds on Mayer's (1997; Mayer, Steinhoff, Bower, & Mars, 1995) generative theory of multimedia learning, which posits that learners engage in three major processes—selecting, organizing, and integrating—when they are presented with visual and verbal information such as illustrations and text. First, when presented with text, the learner must select relevant words to be retained as a text base in verbal working memory, and when presented with illustrations, must select relevant images to be retained as an image base in visual working memory. Second, the learner must organize the text base into a coherent verbal representation and must organize the image base into a coherent visual representation. Third, the learner must integrate the verbal and visual representations by making one-to-one connections between features of the two representations. According to generative theory, meaningful learning is enhanced when a learner can construct and coordinate visual and verbal representations of the same material.

The generative theory of multimedia learning draws on Wittrock's (1974, 1990) generative theory and Paivio's (1971, 1990) dual-coding theory as well as on extensions to both theories, and is further grounded in a number of studies that test the conditions under which the presentation of visual and verbal material promotes learning and for whom multimedia instruction is effective. Mayer and Sims (1994), for example, stated that one of the most important functions of instructional materials (e.g., multimedia software) is to help students construct referential connections between two forms of mental representation: the verbal representational system and the visual representational system. These referential connections are more easily built when both verbal and visual materials are presented continguously. Mayer and Sims's (1994) results are in line with a large body of research that shows that learning can be supported by presenting visual and verbal learning materials in a specific manner (e.g., Mandl & Levin, 1989; Mayer & Anderson, 1991; Rieber, 1990; Schnottz, 1993; Schnottz & Kulhavy, 1994; Winn, 1991) and that multimedia presentations are helpful for certain types of learners (e.g., Mayer & Gallini, 1990; Mayer & Sims, 1994; Mayer et al., 1995).

In extending the generative theory of multimedia learning specifically to multimedia second-language learning, two aspects of language learning are targeted, namely the learning of individual vocabulary words and the overall comprehension of a reading text. Previous studies of the effects of visual associations on memory found that, in the acquisition of vocabulary, foreign words associated with actual objects or imagery techniques are learned more easily (e.g., Kellogg & Howe, 1971). The application of the generative theory of multimedia learning to vocabulary learning suggests, first, that learners of a second language have two separate verbal systems and a common imagery system and, second, that translations of words would not only link the two verbal systems but that this storage in the second verbal system would have an additive effect on learning (cf. also Paivio & Desrochers, 1980; Paivio & Lambert, 1981). Words that are coded dually in two modes (verbally and nonverbally, i.e., with pictures) would be learned better than those coded only verbally.1 Danan (1992), in a study involving the use of video and subtitling, found that a combination of video input and bimodal verbal input (subtitles in the foreign language along with audio in the learner's native language) aided vocabulary learning. This finding was said to support Paivio's dual-coding theory in that "once translation has linked the two verbal systems, students have established more paths for retrieval and can benefit from visual traces as well as from two distinct sets of verbal traces" (p. 522). In other words, students learn new words when they can establish a direct connection between a word in their native language, the corresponding picture of an object or action, and its foreign equivalent. They thus build two types of retrieval cues in memory.

In contrast to vocabulary learning, which involves rote learning on a low level, reading comprehension is a constructive process, involving construction of meaning on a higher level (cf. Mayer, 1984; Wittrock, 1990; Wittrock, Marks, & Doctorow, 1975). Previous foreign language reading research found that visual advance organizers are effective facilitators of comprehension of texts in a foreign language (Levie & Lentz, 1982; Levin, 1981). Omaggio's (1979) study of reading comprehension in French, in which a variety of visual contexts were used as advance organizers, found that students who had still images depicting actions from the story had significantly fewer comprehension errors and recalled a significantly higher number of facts and inferences from the text than those who were provided with only text.

According to the generative theories of comprehension of Mayer (1984) and Wittrock (1990), when reading a text, learners have to build referential connections in working memory between the mental representations of ideas or propositions that have been presented in different modes. Comprehension occurs when these connections are stored in long-term memory, but storage may be hindered if learners are not able to build them. The building of such connections may be affected by individual differences among learners, specifically in the domain of verbalizer-visualizer learning preferences, to be discussed in the following section.

1 It is important to note that our visual–verbal distinction does not follow the strict or traditional distinction between sensorimotor modalities and symbolic (verbal vs. nonverbal) modalities. Mayer (1997), for example, distinguished between presentation modes and sensory modalities. Presentation modes refer to the format used to represent the presented information, (e.g., words vs. pictures); sensory modality refers to the information-processing channel that a learner uses to process the information (e.g., acoustic vs. visual information processing). This article deals not with perception (i.e., sensorimotor modalities) but rather with information processing and, therefore, distinguishes between verbal modes (text translations) and visual modes (nonverbal pictures and videos). In other words, the visual modes used here are a subset of nonverbal modes.
Visualizer and Verbalizer Learning Preferences

One important characteristic of learners is the way they interact with their environment and specifically how they acquire and process visual versus verbal information. Individual differences among learners have been the focus of a large body of studies (cf. Corn & Snow, 1986; Cronbach & Snow, 1977; Denis, 1982; Jonassen & Grabowski, 1993; Keele, 1989; Kirby, 1993; Mayer, 1997). They become an important issue for multimedia learning because this technology allows for the development of adaptive systems that provide learner-controlled options to support the user's preferences, which in turn enhance learning (cf. Cronbach & Snow, 1977; Leutner, 1992, 1995). As an information-gathering style, describing the way students interact with their environment and extract information from it, the visualizer–verbalizer dimension describes individual differences among students when they acquire and process visual versus verbal information. It has also been found to be a good predictor of various task performances in learning with media (Kirby, Moore, & Schofield, 1988; Leutner & Plass, 1997; Schofield & Kirby, 1993). The terms cognitive style and learning style are not used consistently in the literature; thus, for the purpose of this article we refer to the visualizer–verbalizer dimension as learning preference. Also, although recognizing that texts are also perceived visually, we refer, following the visualizer–verbalizer literature, to text information as verbal information and pictures and movies as visual information.

Related to the generative theory of multimedia learning described previously, the potentially moderating effects of visualizer–verbalizer preferences need to be investigated. For example, when both visual and verbal modes of information are present, a general effect of more effective learning can be expected. However, how will learners perform when one mode of information is missing, possibly their preferred mode? Visualizers, for example, might not be able to compensate for the missing mode. In addition, this effect may be stronger in high-order learning (e.g., reading comprehension) than in simple learning of facts (e.g., vocabulary learning).

A small body of research exists on the identification of learning styles or learning preferences among language learners (cf. Corbett & Smith, 1984; Reid, 1987; Reinert, 1976) and on the effectiveness of multimodal materials for students with particular learning styles (cf. Pouwels, 1992). The aforementioned studies, as well as most previous research on learning styles, however, all rely on self-reported questionnaire or inventory data to identify learning styles. Because the psychometric properties of many of these instruments are being questioned (e.g., Boswell & Picket, 1991; Corbett & Smith, 1984; Edwards & Wilkins, 1981; Finke, 1993; Kirby et al., 1988; Parrott, 1986), alternative methods for determining learning preferences are desirable. In this article, visualizer–verbalizer learning preferences of second-language learners are thus identified on the basis of behavioral data (i.e., learners' actual behavior when reading a text in a foreign language; Leutner & Plass, 1997).

Hypotheses

The present article is concerned with a generative theory of multimedia learning and its application to the design of multimedia learning environments. In addition, possibilities of supporting visual and verbal learning preferences to improve learning outcomes are investigated.

With regard to vocabulary learning, the generative theory of multimedia learning suggests that students are more likely to learn the translations of German words when they select and process both verbal and visual information than when they select information in one mode or none. This first hypothesis is based on the idea that students can build two types of retrieval cues when word annotations are presented in verbal and visual forms; in other words, they can integrate the mental representations constructed from the verbal and visual information, whereas they can only build one or no type of retrieval cue when they select annotations in one or no forms. Concerning individual differences in vocabulary learning, it is expected that students’ recall of the translations of German words is better when they report using their preferred mode to retrieve them rather than their unpreferred mode. This second hypothesis is based on the idea that students tend to rely on retrieval routes corresponding to their preferred learning styles, such that visual learners are more likely to benefit from annotations being presented visually and verbal learners are more likely to benefit from annotations being presented verbally.

Unlike vocabulary acquisition, which involves verbal learning on a very low level, reading comprehension requires verbal learning on a higher level, including understanding words in context as well as propositions and thus constructing meaning. The understanding of words and propositions may depend on the availability of the preferred mode of information for a particular learner. The effect of learning from verbal and visual presentation modes (by organizing text and images into coherent mental representations and integrating these representations with one another) would be moderated by individual differences in learning preferences in that the unavailability of the preferred mode of information would limit the understanding of propositions in which a word is embedded and could hinder its storage in memory altogether. Thus, concerning individual differences in reading comprehension, as a third hypothesis, visualizers are expected to recall propositions that contain words with visual and verbal annotations better than propositions that contain words with only verbal annotations. For verbalizers, this difference is expected to be smaller because they receive their preferred mode a priori. Also concerning individual differences in reading comprehension, as a fourth hypothesis, visualizers are expected to recall text propositions that were additionally visually illustrated in a preview video better than propositions that were not. For verbalizers, this difference is again expected to be smaller. The latter two hypotheses are based on the idea that visualizers tend to rely on visual representations for building a mental model of the situation being described in the text, whereas verbalizers tend to rely on verbal representations.
Method

Participants and Design

The participants were 103 college students who were enrolled in second-year German language courses at a highly selective university in California. All students were nonnative speakers of German and were fluent in English. Students participated in the study as a regular class activity; thus, they were in a real learning situation in an intact class setting, being asked to learn from authentic materials. The mean grade point average of the students was 3.42 (SD = 0.41; with A = 4.0, B = 3.0, C = 2.0, D = 1.0, and F = 0). On the basis of an analysis of preferences for visual and verbal information in a multimedia learning task as described in the procedure section, 39 students were classified as visualizers, 35 were classified as verbalizers, and 29 were classified as showing no strong preferences. All students took the same vocabulary test and comprehension test.

Materials and Apparatus

The apparatus for developing the multimedia program consisted of a Macintosh Centris 650 with a Rasterops MoviePak board for video digitizing. The apparatus for presenting the multimedia program to students consisted of a 20-station computer lab equipped with Macintosh LC II computer systems with 12-inch color monitors, 40-megabyte hard disks, 4 megabytes of RAM, and headphones.

The materials consisted of a multimedia program that ran on Macintosh color computer systems, including a computer-generated preview video. The preview was a 2-min noninteractive video that summarized the key events in the story and served as an advance organizer. It was accompanied by a voiceover in English: “Once upon a time there was a fisherman, and a tourist. The tourist said, "Arbeit [work]." The fisherman replied, "Freizeit [free time].” The story program presented an interactive multimedia version of Heinrich Böll's (1986) short story, *Anekdote zur Senkung der Arbeitsmoral* [Anecdote Concerning the Lowering of Productivity], developed by Chun and Plass (1995, 1997). The story consisted of 762 words in German presented in 11 pages. Each page consisted of approximately 50 to 100 words of text in German presented on the right side of the screen. At the bottom of each page was a right-facing arrow and a left-facing arrow; by using the mouse to click on the right-facing or left-facing arrow symbol, a student could turn to the next page or the previous page, respectively. Several of the words on each page were marked by a degree symbol, with 82 words marked overall (as shown in Figure 1). If a student moved the mouse to one of these words and pressed the mouse button, one or more option symbols would appear at the top of the screen (i.e., icons representing text annotations, pictures, and videos). For each of the marked words, a text translation was available; for some of them a picture or video was available in addition. To select an option, the student was instructed to drag the word to the corresponding icon. When the translation option was selected, the computer presented the word spoken by a native speaker of German via the headphones and presented a textual translation in English in a window on the left side of the screen. For example, for the German word “Fischschwärme,” the translation was “schools of fish.” When the picture option was selected, the computer presented the word spoken by a native speaker of German and presented a picture corresponding to the meaning of the word in a window on the left side of the screen. For example, for the German word “Fischschwärme,” the picture showed a school of fish. When the video option was selected, the computer presented the word spoken by a native speaker of German and presented a short video corresponding to the meaning of the word in a window on the left side of the screen. For example, for the German word “Hubschrauber,” the video consisted of a 7-s video clip of a helicopter. To close the window with the selected information and return to reading the story, the student was instructed to click on the mouse within the window. The options could be selected any number of times and in any order. The program recorded chronologically every action of the student, including each selected option and each turned page. Figure 1
shows a page from the story program. The English translation of this portion of the story is: "[awakens the dozing] fisherman, who sleepily sits up, sleepily gropes for his cigarettes, but before he has found what he is looking for the eager tourist is already holding a pack under his nose, not exactly sticking a cigarette between his lips but putting one into his hand, and a fourth click, that of the lighter, completes the overeager courtesy. As a result of that excess of nimble courtesy—scarcely measurable, never verifiable—a certain awkwardness has arisen that the tourist, who speaks the language of the country, tries to bridge by striking up a conversation."

The program was written in HyperCard (Apple Computer, Inc., 1993), the audio was processed with Sound Designer II (Digidesign, 1992) and Sound Edit Pro (Macromedia, 1991), the pictures were processed with Adobe Photoshop (Adobe, 1993), and the video clips were digitized with MoviePak (Rasterops, 1993) and processed with Adobe Premiere (Adobe, 1994).

The pencil-and-paper materials consisted of a questionnaire, vocabulary posttest, and comprehension posttest each typed on 8.5 × 11-in. sheets of paper. The questionnaire solicited information concerning each participant’s gender, native language, grade point average, and grades in previous German classes. The vocabulary posttest included a list of 24 words with both visual and verbal annotations.

Twelve of the 24 marked words had text and picture options: sprengen (to explode), dösen (to doze), aufdauen (to thaw), drohen (to threaten), Fang (catch), Fischschwärme (schools of fish), Kutter (cutting), Fischermütze (fisherman’s cap), ärmlich (poor, poorly), maßbar (measurable), besorgt (anxious), and nachdenklich (pensive). The other 12 words had text and video options: sich aufrichten (to sit up), Kopfschütteln (shaking of one’s head), Zeichensprache (sign language), sich recken (to stretch), Feuerschwein (lighter), Wellenkämme (whitecaps), Hubschrauber (helicopter), Münzer (mouths), gereizt (irritated), betrübt (sad), verpaßt (missed), and feindselig (hostile). Students were instructed to produce an English translation of each word. In addition, for each word students were instructed to check one of four boxes to indicate whether the German word primarily reminded them of hearing the word pronounced, reading a printed translation, seeing a picture, or seeing a video. The comprehension test was a sheet of paper along with instructions to write a summary of the story or “recall protocol” in English.

Procedure

Students were tested in their intact German language classes in groups of 15 to 20 per session during their normally scheduled class hour. Classes met for 50 min per day, and the entire procedure required two 50-min class periods on 2 consecutive days. On the 1st day, students first filled out the pencil-and-paper questionnaire at their own rates. Each student was seated in front of a separate Macintosh computer system in a language computer lab. Second, after collection of the questionnaires, students were given a brief demonstration of the program on a large television monitor at the front of the lab. Students were told how to turn pages and how to look up a marked word. The instructor demonstrated how to click on a marked word, hold the mouse button down, and, when the icons appeared indicating which type of annotation was available, how to drag the word to the desired icon and release the mouse button. The students were told that, on selecting an option, they would hear the word pronounced in German and see either the written translation, sound, picture, or movie, and whether the word was known before reading the story. In constructing the comprehension test, two raters independently listed the main idea unit or propositions of the text, identified the main idea and occurrences in the passage. Each rater originally came up with 10 propositions, agreeing on 9 of the 10, and through discussion came to a consensus about the 10th and 2 additional ones (Appendix). Each recall protocol was then scored in terms of these 12 propositions: 1 point was given for each of the propositions that was mentioned, and totals for each student as well as for each of the

In addition, two pencil-and-paper tests were administered to identify verbalizers and visualizers: the Visualizer/Verbalizer Questionnaire by Kirby et al. (1988) and the Edmonds Learning Style Identification Exercise by Reintert (1976). The results of the tests were not used in this study because of the low level of agreement between the two tests as well as other reported problems with the reliability of the tests (Boswell & Picket, 1991; Corbett & Smith, 1984; Edwards & Wilkins, 1981; Keefe, 1989; Parrot, 1986).

Because an authentic story was used that did not allow for visual depictions of all words, it was not feasible to balance the vocabulary words used for the posttest after a completely crossed factorial design involving factors that could affect the internal validity of the results of the study (e.g., word type, concreteness, difficulty, significance, and occurrence). However, every effort was made to ensure that the words in each of the categories (picture + translation and video + translation) were equivalent in terms of these five different factors: First, in terms of word type, there were equal numbers of words for actions, objects, and descriptions. Second, within word types, there were equivalent numbers of words for human versus nonhuman actions and for human versus nonhuman descriptions. Third, in terms of difficulty of words, the words in each category were equivalent based on ratings by experienced German instructors. Fourth, in terms of significance of the words for the story, the words in each category were equivalent based on whether or not they were included in one of the 12 major propositions of the story. Fifth, the words in each category were also equivalent based on where (on what page) the words occurred in the text.

Students had been told on the 1st day to return on the 2nd day to fill out some questionnaires and do some “exercises.” They were told explicitly what types of “exercises” to expect, and were given neither written copies of the short story nor access to the computer lab to use the multimedia computer program. They, therefore, had no opportunities to review what they had learned the previous day or to “rehearse” any of the information.
individual propositions were tallied (cf. Lee & Ballman, 1987, regarding determining idea units and scoring, and Deville & Chalhoub-Deville, 1993, who found no difference in scores whether recalls are scored dichotomously or are weighted). The internal consistency of this measure is \( \alpha = 0.85 \).

Identification of learning preferences. For the identification of learning preferences, the log-file data were used. These files were recorded by the program while students read the text and contain a record of the type of annotations students looked up for each word. For all of the 24 words that were annotated with both visual and verbal information and that were looked up by an individual student, the frequency of choosing a verbal annotation only, or, if both visual and verbal annotations were selected, the frequency of looking up a verbal annotation first was calculated. Similarly, the frequency of choosing a visual annotation only or looking up a visual annotation first was determined. The difference between these two measures, divided by the total number of words looked up, determined each student's position on a bipolar scale, with extreme visual preferences on one end and extreme verbal preferences on the other end. Thirty-five students scored above the 66th percentile and were classified as verbalizers; 29 scored below the 33rd percentile and were classified as visualizers; 29 scored between the 33rd and 66th percentiles and were classified as controls. To examine the reliability of this measure of learning preference, students' preference scores for words annotated with pictures and text were correlated with their preference scores observed for words annotated with video and text. The resulting correlation of .72 turned out to be high and significant \((p < .05)\), indicating high reliability of the test scores.

Results

Hypothesis 1: Students Are More Likely to Learn the Translations of German Words When They Are Exposed to Verbal and Visual Instruction Than When They Receive Instruction in One Mode or None

According to the generative theory of multimedia learning, the acquisition of a specific vocabulary word while reading a story is best when both verbal and visual information is looked up, because when a student has actively processed both visual and verbal annotations of a word and integrated the constructed visual and verbal mental representations with one another, he or she is more likely to have two kinds of retrieval routes when presented with the target word to be defined on a test rather than one or no retrieval route. This multimedia retrieval-route analysis of rote vocabulary learning leads to the prediction that, if all other conditions are equal, vocabulary test performance will be worst when no information at all had been looked up during learning, moderate when one type of information had been looked up, and best when both visual and verbal information had been looked up.

When reading the story, each student had complete freedom to look up visual (i.e., picture or video), verbal (i.e., text translation), or both annotations of specific words they did not know, given that both kinds of annotation were available for that word. Because a student's look-up behavior may change from word to word, the only way to test the hypothesis is to use vocabulary items, not students, as the units of observation. In this case, for each vocabulary item, data are thus aggregated over students. That is, for each of the 24 items on the vocabulary test that were annotated in the story with both verbal and visual options, each of the 103 students was classified according to his or her look-up behavior with regard to that specific item while reading the story. There are four such classes of look-up behavior: (a) the student did not look up any annotation (none group), (b) the student looked up the visual annotation only (visual-only group), (c) the student looked up the verbal annotation only (verbal-only group), or (d) the student looked up both the visual and the verbal annotation (visual and verbal group).

For each of the 24 words in the vocabulary test, every student was assigned to one of the four groups according to his or her look-up behavior for this word. Then, for each word, and for each of the four groups of students, the percentage of correct answers was determined, yielding a \( 24 \times 4 \) data matrix (24 test items as units of analysis with four measures for each item).

---

5 This extreme group definition identifies individuals as visualizers and verbalizers who are not pure cases of their classes. Although this is a traditional approach for analyzing the influences of individual differences, it is acknowledged that the statistical power of such an approach is reduced compared with an approach in which the individual differences variable is treated as a metric variable (Maxwell & Delany, 1993; Leutner & Rammsayer, 1995). The results of all analyses concerning Attribute × Treatment interaction effects in this article, however, were verified according to Alexander and DeShon (1994) by testing the difference between corresponding correlations using the whole sample. The results show that in each case in which a significant Attribute × Treatment effect was found with extreme groups \((n = 74)\), this effect was also found by testing correlations in the whole sample \((N = 103)\).

6 Unlike questionnaire-based methods to identify learning preferences, this instrument uses the observation of preferential choice behavior in an authentic learning situation and thus provides a direct measure of the strength and consistency of the learner's preference for visual versus verbal learning material. Cronbach's alpha is .92. For more information about the psychometric properties of this instrument, see Leutner and Plass (1997).

7 Note that for each word a student can be in different groups, depending on his or her look-up behavior, which can vary from word to word. Across all 24 words, the average number of students per group was 23.9 \( (SD = 25.8; \text{none group}) \), 29.0 \( (SD = 25.5; \text{visual-only group}) \), 33.5 \( (SD = 28.3; \text{verbal-only group}) \), and 40.4 \( (SD = 30.0; \text{visual and verbal group}) \). The difference between these group sizes is not statistically significant. \( F(3, 69) = 2.03, \text{MSE} = 184.65 \), which can, in conjunction with the high internal consistency of the instrument \((\alpha = .92)\), be interpreted as an indication against a novelty effect, in which students choose a certain type of information out of curiosity rather than on the basis of their preferences. This may be because the study was conducted as part of the curriculum and was thus taken very seriously by the participants. Also, students were in fact allowed "double exposure" (i.e., they could choose to see any annotation multiple times). However, because of the authentic character of the learning situation, the learners were not forced to look up anything twice. Log files showed that very few students (average across words: 4.2%) actually did select the same annotation more than once. Thus, double exposure within one mode can be neglected for this study.
Figure 2 shows the mean percentage of correct answers to vocabulary test items for which both visual and verbal options were available broken down by students who chose no option, a visual option, a verbal option, and both a visual and verbal option. An analysis of variance (ANOVA) with look-up behavior group as a repeated measures factor revealed that the differences are significant, $F(3, 69) = 7.08$, $MSE = 166.78$, $p < .001$. Supplemental pairwise comparisons (one-tailed test) revealed that, compared with looking up no information, it was helpful to look up visual information ($t = 2.95$, $p = .004$, $SE = 3.27$), verbal information ($t = 3.62$, $p < .001$, $SE = 4.57$), although only marginally significant for visual information. As a trend, visual information was found to have helped less than verbal information ($t = 1.38$, $p = .091$, $SE = 3.74$), verbal information ($t = 2.95$, $p = .004$, $SE = 3.27$), or both ($t = 3.62$, $p < .001$, $SE = 4.57$), although only marginally significant for visual information. As a trend, visual information was found to have helped less than verbal information ($t = 1.32$, $p = .099$, $SE = 3.37$). Finally, looking up both types of information resulted in significantly better results than looking up either visual information only ($t = 3.43$, $p < .001$, $SE = 3.32$) or verbal information only ($t = 1.76$, $p = .046$, $SE = 3.96$).

As predicted, there is a strong effect of students' look-up behavior on vocabulary acquisition such that learning the translation of foreign words is best when students selected both visual and verbal modes of instruction, moderate when they selected only one mode of instruction, and worst when they selected neither. This result extends earlier work on active learning strategies by demonstrating that the combination of visual and verbal learning modes results in higher performance than one mode alone.

As predicted, there is a strong effect of students' look-up behavior on vocabulary acquisition such that learning the translation of foreign words is best when students selected both visual and verbal modes of instruction, moderate when they selected only one mode of instruction, and worst when they selected neither. This result extends earlier work on active learning strategies by demonstrating that the combination of visual and verbal learning modes results in higher performance than one mode alone.

Hypothesis 2: Students'Recall of the Translations of German Words Is Better When They Report Using Their Preferred Mode to Retrieve Them Rather Than Their Unpreferred Mode

According to the foregoing analysis, students perform better in retrieving the translation of a German word when they choose to study both visual and verbal annotations because they have two different types of retrieval cues from the word to its meaning. This second hypothesis concerns individual differences between students who prefer to learn from visual annotations (i.e., those who tend to first select a visual option when they come to a marked word with both visual and verbal options) and those who prefer to learn from verbal annotations (i.e., those who tend to first select a verbal option when they come to a marked word with both visual and verbal options).

On the basis of the retrieval-cue analysis of the generative theory of multimedia learning, we predict that, when giving a correct answer on the vocabulary test to 1 of the 24 words that allowed for both visual and verbal look-up options, verbalizers will report being reminded of verbal information more often than will visualizers, whereas visualizers will report being reminded of visual information more often than will verbalizers. However, no differences between visualizers and verbalizers in the overall percentage of correct answers on the vocabulary test are expected.

Figure 3 shows the mean percentage of correct responses on the vocabulary test for items on which visualizers and verbalizers reported remembering translations via visual cues (such as a picture or video) and for items on which visualizers and verbalizers reported remembering translations based on verbal cues (such as a text translation). The
Better than propositions that contain words with comprehension test by visualizers and verbalizers for propositions that allow for both visual and verbal modes of learning. When students have access to their preferred learning mode, they remember more than those who reported remembering visual annotations. Verbalizers performed significantly better on words that reminded them of verbal annotations than words that reminded them of visual annotations, whereas verbalizers showed the opposite pattern.

The results suggest that visualizers are more effective in using visual cues for remembering vocabulary information, whereas verbalizers are more effective in using verbal cues. An instructional implication is that multimedia environments that allow for both visual and verbal modes of elaborating on words may be effective because learners can choose the mode of annotation that best suits their learning preference.

Hypothesis 3: Visualizers Recall Propositions That Contain Words With Visual and Verbal Annotations Better Than Propositions That Contain Words With Only Verbal Annotations, Whereas for Verbalizers This Difference in Recall Between the Two Types of Propositions Is Smaller

To comprehend the story globally, students had to understand the main propositions that make up the story. These propositions contain words, some of which may be unknown to the reader, so that looking up certain unknown words may be a prerequisite for comprehension. Some of the words that were not part of the 24-item vocabulary test allowed only for a verbal annotation, and some of the words allowed for both verbal and visual annotation. We predict a pattern of Attribute × Treatment interaction in which students perform better when they have access to the multimedia formats that are consistent with their preferred modes of learning than when they do not.

Figure 4 shows the mean percentage correct on the comprehension test by visualizers and verbalizers for propositions that allowed for verbal annotation only and propositions that allowed for both verbal and visual annotation. As predicted, there is a pattern of Attribute × Treatment interaction in which visualizers perform well in recalling both types of propositions, whereas verbalizers perform well on propositions containing both verbal and visual annotation options but not on those containing only verbal annotation options. Presumably, verbalizers had access to their preferred learning mode in both types of propositions, whereas visualizers had access to their preferred learning mode only in the visually and verbally annotated propositions. Thus, when students are given the opportunity for active learning in their preferred learning mode, they remember more than when they are not.

To examine these observations, we conducted an ANOVA on the data in Figure 4, with learning preference (visualizers vs. verbalizers) as a between-subjects factor and type of proposition (verbally annotated vs. visually and verbally annotated) as a within-subject factor. There was no significant main effect of learning preference, \( F(1, 72) = 2.55, \text{MSE} = 79.40, \) confirming the expectation that the overall performance on generating translations between visualizers and verbalizers was not different. There was also no significant main effect of annotation type, \( F(1, 72) < 1, \text{MSE} = 90.40, \) confirming the expectation that students who reported remembering verbal annotations did not remember more translations overall than did students who reported remembering visual annotations. The major focus of the analysis revealed a significant interaction of learning preferences and annotation type, \( F(1, 72) = 45.9, \text{MSE} = 90.40, p < .001, \) in which visualizers performed significantly better on words that reminded them of visual annotations than words that reminded them of verbal annotations, whereas verbalizers showed the opposite pattern.

The results suggest that visualizers are more effective in using visual cues for remembering vocabulary information, whereas verbalizers are more effective in using verbal cues. An instructional implication is that multimedia environments that allow for both visual and verbal modes of elaborating on words may be effective because learners can choose the mode of annotation that best suits their learning preference.

**Figure 4.** Percentage of correct answers in the comprehension test with respect to type of annotation available (verbal, visual) and learning preference (visualizer, verbalizer).
types of propositions. Consistent with predictions, the interaction was statistically significant, $F(1, 72) = 4.42$, $MSE = 429.20, p < .05$.

**Hypothesis 4: Visualizers Recall Text Propositions That Were Visually Illustrated in the Preview Better Than Propositions That Were Not, Whereas for Verbalizers This Difference in Recall Between the Two Types of Propositions Is Smaller**

In the foregoing analysis involving the comprehension test, the positive effect of having access to both visual and verbal modes of information was present for visualizers but not for verbalizers. Thus, for comprehension it seems that verbalizers do not need visual information in addition to their preferred verbal type of information, whereas visualizers perform worse when their preferred type of information is not available. This interesting finding warrants a replication with a different set of data. Therefore, in this analysis, we examine the effect of the visual preview on the comprehension of visualizers and verbalizers. To validate the results of the previous analysis, it was expected that visualizers would recall propositions that were illustrated in the preview video better than propositions that were not illustrated, whereas a smaller difference was expected for verbalizers.

Figure 5 shows the mean percentage correct on the comprehension test for visualizers and verbalizers on propositions that had been visually illustrated in the preview and those that had not. Consistent with predictions, there is a pattern in which visualizers recalled the illustrated propositions much better than the unillustrated propositions, whereas the verbalizers recalled the illustrated propositions only slightly better than the unillustrated propositions.

An ANOVA on the data summarized in Figure 5 was conducted with learning preference (visualizers vs. verbalizers) and proposition type (five propositions that had been illustrated vs. five propositions that had not been illustrated) as factors. There was no significant effect of learning preference, $F(1, 72) < 1$, but there was a statistically significant effect of proposition type, $F(1, 72) = 45.46$, $MSE = 228.50, p < .001$, in which students recalled illustrated propositions better than unillustrated propositions. The main focus of this analysis was on the interaction of learning preferences and preview illustration, which was statistically significant, $F(1, 72) = 5.06, MSE = 228.50, p < .05$. The significant interaction supports the prediction that the effect of providing a visual preview of a proposition is larger for visualizers than for verbalizers.

The results of this analysis replicate the highly interesting Attribute × Treatment interaction effect of the previous analysis. Consistent with previous results by Mayer and Sims (1994), the results for the comprehension test reveal an interaction in which verbalizers do not profit from the addition of visual information to their preferred verbal type of information to the same extent as visualizers do. Reinterpreted from the perspective of visualizers, this means that visualizers perform worse when their preferred type of information is not available when reading and trying to comprehend a literary text.

**Discussion**

A basic theme of this study is that students learn more effectively when they have access to and actively select visual and verbal modes of elaborating on presented material than when they have access to or select only one mode or neither. In addition, a secondary theme is that visualizers benefit from visual modes of elaborating on the material more than do verbalizers, whereas verbalizers benefit more from verbal modes of elaborating on presented material than do visualizers. The findings emphasize the importance of individual differences, such as learning preferences in the visualizer–verbalizer dimension, for the study of learning with media.

The results concerning vocabulary acquisition are in line with the generative theory of multimedia learning. Students’ performance on the posttests was best when both visual and verbal modes of instruction were selected, moderate when students selected only one mode of instruction, and worst when they selected neither. In the vocabulary posttest, visualizers were more likely to produce a translation correctly when they reported using a visual retrieval cue (i.e., being reminded of a corresponding picture or video) than when they reported using a verbal retrieval cue (i.e., being reminded of reading a text translation), whereas verbalizers were more likely to produce a translation correctly when they reported using a verbal retrieval cue rather than a visual one.

For reading comprehension, however, an Attribute × Treatment interaction was found between learning prefer-
ence and type of information used to annotate a proposition; visualizers performed better on propositions that allowed for visual and verbal annotations than on those that allowed only for verbal annotations, and verbalizers performed well on both types of propositions. This Attribute × Treatment interaction was replicated using a preview video that visually illustrated some of the propositions of the text. Visualizers performed better on propositions that were illustrated in the preview than on those that were not, and verbalizers performed well on both types of propositions.

This study offers educational and theoretical implications. On the educational side, it provides evidence relevant to the design of multimedia instruction in general and for second-language learning in particular. Specifically, this study suggests that learners should have options for selecting and processing material presented in both visual and verbal modes. In our study, students learned vocabulary words best when they selected both visual and verbal annotations, and they recalled propositions best when they had the option of selecting both visual and verbal annotations and when the propositions were illustrated in a preview video. In addition, providing both options is most effective in addressing individual differences in visual and verbal learning preferences. Concerning second-language learning, this study points to the advantages of adding a visual component to the traditionally strong emphasis on verbal modes of instruction.

On the theoretical side, this study provides evidence for the generative theory of multimedia learning (Mayer, 1997), which makes a distinction between building visual and verbal cues for retrieving stored information from memory in vocabulary learning and building visual–verbal referential connections for storing the global comprehension of a story in memory. Vocabulary learning, on the one hand, can be viewed as rote learning. The information to be stored in memory is explicitly given to the student, and the storage process itself should not be a major problem. Retrieving the stored information, however, can be difficult. This is the situation in which the organization and integration of two different forms of mental representations enhance retrieval performance by providing multiple retrieval cues. Interestingly, our data suggest that individual differences on the visualizer–verbalizer dimension play only a minor role in retrieval performance.

For reading comprehension, the situation is quite different. The information to be stored in memory is not explicitly given to the student but has to be constructed by him- or herself when reading a text. Thus, although storage should not be a major problem in vocabulary learning, it may be a severe problem for reading comprehension, specifically for those students who are unable to build an adequate mental model of the text. This is the point at which individual differences may come into play. It can be expected, for example, that visual learners, when their preferred mode of processing information is not present while reading a foreign language text, will be unable to make the visual and verbal referential connections necessary for comprehension, whereas verbal learners' comprehension will not be affected to the same extent when visual information is absent because their preferred verbal mode is sufficient. Thus, for reading comprehension of a foreign language text, it can be argued that the effect of selecting and processing both visual and verbal modes of information is moderated by individual differences in performance during the storage process of constructed meaning. Our data provide strong evidence for such an extension of the generative theory of multimedia learning to second-language learning.

In addition, results of this study concerning the nature of individual differences in visual and verbal learning preferences suggest a new method of assessing these learning preferences. Visualizers tend to benefit from visual activities during learning, whereas verbalizers tend to benefit from verbal activities during learning. This preferential choice behavior in a learning situation offers a method to identify visualizer–verbalizer learning preferences with better psychometric properties compared with most conventional self-report questionnaires (Leutner & Plass, 1997).

References


Appendix

Major Propositions

Anekdote zur Senkung der Arbeitsmoral
(Ancedote Concerning the Lowering of Productivity)

by Heinrich Böll

1. A fisherman is lying in his boat dozing.
2. A tourist comes and starts photographing him.
3. The photographing wakes (and annoys) the fisherman.
4. The tourist tries to engage the fisherman in conversation.
5. The tourist asks about the fishing that day and whether the fisherman is going out again.
6. The fisherman replies he had a good day and caught enough for the next several days.
7. The tourist starts telling the fisherman what would happen if the fisherman would go out more often.
8. The tourist speculates about all sorts of things the fisherman could have and do if he “worked harder.”
9. The tourist says the ultimate result would be that the fisherman could lie in the sun and doze all day.
10. The fisherman replies that he is already doing that.
11. The tourist leaves, envying the fisherman for that.
12. The “moral of the story” is . . . .

Received May 17, 1996
Revision received August 8, 1997
Accepted August 9, 1997

Low Publication Prices for APA Members and Affiliates

Keeping you up-to-date. All APA Fellows, Members, Associates, and Student Affiliates receive—as part of their annual dues—subscriptions to the American Psychologist and APA Monitor. High School Teacher and International Affiliates receive subscriptions to the APA Monitor, and they may subscribe to the American Psychologist at a significantly reduced rate. In addition, all Members and Student Affiliates are eligible for savings of up to 60% (plus a journal credit) on all other APA journals, as well as significant discounts on subscriptions from cooperating societies and publishers (e.g., the American Association for Counseling and Development, Academic Press, and Human Sciences Press).

Essential resources. APA members and affiliates receive special rates for purchases of APA books, including the Publication Manual of the American Psychological Association, and on dozens of new topical books each year.

Other benefits of membership. Membership in APA also provides eligibility for competitive insurance plans, continuing education programs, reduced APA convention fees, and specialty divisions.

More information. Write to American Psychological Association, Membership Services, 750 First Street, NE, Washington, DC 20002-4242.