The Function of Annotations in the Comprehension of Scientific Texts: Cognitive Load Effects and the Impact of Verbal Ability

□ Erik Wallen
Jan L. Plass
Roland Brünken

Students participated in a study (n = 98) investigating the effectiveness of three types of annotations on three learning outcome measures. The annotations were designed to support the cognitive processes in the comprehension of scientific texts, with a function to aid either the process of selecting relevant information, organizing the information in memory, or integrating information with prior knowledge. Learning outcomes were measured by assessing student recall of facts, comprehension of the text, and mental model construction. Results show that different types of annotations facilitate different learning outcomes. In addition, we found that, compared to having only one type of annotation available, multiple types of annotations resulted in a higher cognitive load that resulted in lower performance, especially in tests of higher-level processing. This effect was stronger for low-verbal-ability learners, who showed lower performance in treatments with multiple types of annotations than high-verbal-ability learners.

□ How can Web-based instruction and multimedia learning environments be designed to aid learners in the comprehension of science materials? Consider a scenario in which learners read text on a Web page that describes how cellular phones work. As they read the description, the learners encounter certain words or concepts that they are unfamiliar with. Unlike a book, however, the computer-based nature of the presentation of the text allows them to click on words or phrases to receive definitions or descriptions of the unfamiliar concepts. Using these annotations, they are able to understand the concept in question, and thereby comprehend a text that might otherwise have been too complex for their level of prior knowledge.

To aid the learner in the processing of difficult material, instructional designers often add multimedia annotations to reading texts. Yet little theory-based guidance is available to systematically design such annotations. Research has found that the availability of annotations generally leads to improved learning (Chun & Plass, 1996; Jones & Plass, 2002; Levin, Anglin, & Carney, 1987; Plass, Chun, Mayer, & Leutner, 1998). However, other research has found that adding supplemental information can increase learners' cognitive load and may thus result in reduced learning (Brünken, Plass, & Leutner, 2004; Plass, Chun, Mayer, & Leutner, 2003; Sweller, 1999). In this paper, we investigate the impact of annotations on learner comprehension of textual information of scientific content, the cognitive load caused by multiple types of annotations, and the impact of verbal ability on learning with annotations.

The Function of Annotations in Text Comprehension

In order to provide a systematic approach to the design of annotations for scientific texts, it is necessary to determine what function annotations have in the process of text comprehension. Levin et al. (1987) reviewed the research literature on the use of pictures to supplement prose and categorized the illustrations used in the research into five types: (a) decorative, (b) representative, (c) organizing, (d) interpreting, and (e) transforming. Decorative pictures have a motivational function but have little relation to the content of the prose. Representative pictures depict the content of the instruction, and organizing pictures reveal existing knowledge structures. Interpreting pictures are used to visualize abstract concepts, and transforming pictures support higher-level cognitive processes (Levin et al.). While this post hoc approach has been useful to categorize existing research, which the authors demonstrated by reviewing more recently published articles in a follow-up study (Carney & Levin, 2002), a more systematic approach is desirable to guide the design of instructional materials involving different types of annotations.

In this research, we are extending the definition of functions of picture annotations to also include text annotations, which were the focus of the research reported in this paper.

Cognitive Theory of Multimedia Learning

The cognitive theory of multimedia learning (Mayer, 2001), combining Paivio's (1986) research on dual coding and Wittrock's (1990) research on generative learning, proposed that learning from text and pictures involves three types of cognitive processes. (a) Learners first have to select relevant visual and verbal information from what is presented. (b) Then they must organize this information in meaningful visual and verbal mental representations by building connections within the verbal and the visual systems, and (c) finally integrate the visual and verbal mental representations by establishing connections between the two systems. These cognitive processes are executed under the capacity limitations of working memory (Sweller, 1999).

It has been previously argued that multimedia annotations should be designed to support each of these levels of processing (Chun & Plass, 1997). We now extend this argument to state that each annotation should have a clearly defined function in support of the cognitive processes of text comprehension, namely in support of selecting, organizing, or integrating the information. Such a function would allow the instructional designer to systematically construct annotations based on an understanding of their cognitive impact.

The Cognitive Function of Annotations in Text Comprehension

We have defined three types of cognitive functions of annotations in text comprehension based on Mayer's (2001) theory of multimedia learning; a comparison to the five categories put forth by Levin et al. (1987) reveals striking parallels. The decorative function, involving pictures with no instructional purpose, is not included as a function in our model. *Decorative* pictures can be interpreted in terms of a coherence effect in Mayer's (2001) model, representing irrelevant pictures that divert attention from important materials. Similarly, based on Sweller's (1999) cognitive load theory, they can be described as extraneous information. *Representative* pictures, defined as depicting the content of the instruction, are similar to selection-level annotations, that is, annotations that support the selection of relevant information. Such a support could be as simple as defining the annotated term, visualizing it, or translating it from a foreign language into the learner's first language. *Organizing* pictures, defined by Levin et al. (1987) as revealing existing knowledge structures, are related to organization-level annotations, that is, annotations that support the learner in organizing the mental representations in working memory into coherent structures. This process combines words into propositions, and fragments of images into a more complete visual representation. *Interpreting* pictures, defined as visualizing abstract concepts, and *transforming* pictures, defined as supporting higher-level cognitive processes, can both be thought of as integration-level annotations, that is, annotations that support higher-level processes, such as the integration of visual and verbal representations, and their integration with prior knowledge.

The availability of selection-level, organization-level, and integration-level annotations, each supporting different cognitive processes, addresses different levels of encoding of information. Selection-level annotations, though enabling the higher-level processes, primarily support encoding of factual knowledge at a low (rote learning) level. Organization-level annotations, supporting the connection of words into ideas, support the comprehension of the information in the text. Integration-level annotations, focusing on relationships between visual and verbal information and prior knowledge, support higher-level cognitive outcomes, such as synthesis, evaluation, and the creation of new ideas (Anderson & Krathwohl, 2001).

The focus of this study was to ascertain if each of the processes of selecting, organizing, and integrating information can be selectively supported by annotations in a science text, and if the availability of these annotations will differentially affect cognitive outcome measures on three different levels: (a) factual knowledge, (b) comprehension, and (c) higher-level learning (synthesis, evaluation, creating).

Cognitive Load

Recent research on multimedia learning has shown that in order to assure the instructional effectiveness of an intervention, cognitive load implications of the design of the materials have to be taken into consideration. Cognitive load theory (Brünken, Plass, & Leutner, 2003; Paas, Renkl, & Sweller, 2003; Sweller, 1999; Sweller, van Merriënboer, & Paas, 1998) describes different sources of working memory load, related to the complexity of the material (intrinsic load), the instructional design (extraneous load), and the amount of mental effort learners invest in learning the materials (germane load). In the case of adding supplemental information such as annotations, the resulting increase in cognitive load could be germane, manifested in improved learning. Although not intended by the instructional designer, the resulting increase in load could also be extraneous, leading to decreased learning. This would be the case, for example, when the information presented is redundant for learners (they already know it), or when the intrinsic load of the materials and the added load by the annotations exceed the cognitive capacity of the learners (Chandler & Sweller, 1991; Sweller, Chandler, Tierner, & Cooper, 1990). Such working memory-related effects are more likely to occur either when multiple types of annotations are available, or when the learner has a low verbal ability (Plass et al., 2003).

Asecond focus of this study is, therefore, if multiple types of annotations will lead to reduced learning, which would be an indication of increased cognitive load, and what impact verbal ability has on the instructional effectiveness of the annotations.

Research Questions and Hypotheses

The present study focused on three questions:

1. Will different types of annotations result in differential learning outcomes in comprehending a scientific text?

   We expect that annotations can be designed with a specific function to support text comprehension on a higher or lower level.

2. Will the availability of two different types of annotations result in increased or decreased learning compared with one type of annotation?

   We expect that the cognitive load generated by two types of annotations will result in decreased learning because of the difficulty of the material and the low prior knowledge of the learners.

3. Does verbal ability have a moderating effect on the load experienced in learning with multiple annotations?

   We expect that low-verbal-ability learners will experience higher load in learning with multiple annotations than will high-verbal-ability learners.

METHOD

Participants

The participants were 109 college students (mean age 19.6 years) recruited from the psychology subject pool at a large public university in southwestern United States. Participants received course credit for their participation. All participants took a prior knowledge test and those with high prior knowledge of physics, electrical engineering, or cell phone systems were excluded from the analysis. Other students had to be excluded from the analysis because of corrupted data files and incomplete data. The remaining 98 participants were randomly assigned to the treatment conditions. Of these, 14 participants served in the no-annotations control group (NA), 14 in the group that received selection-level annotations (S), 15 in the group receiving organization-level annotations (O), and 14 in the group receiving integration-level annotations (I). Ten participants served in the group receiving selection- and organization-level annotations (SO), 17 in the group receiving selection- and integration-level annotations (SI), and 14 in the group receiving organization- and integration-level annotations (OI). The unequal groups sizes are because of the students who had to be excluded from the analysis.

Materials and Apparatus

The computer-based materials consisted of a participant questionnaire; a survey of prior knowledge on cell phones, physics, and electrical engineering; the instructional text; and three tests. The tests were designed to measure different levels of learning according to Bloom's revised taxonomy (Anderson & Krathwohl, 2001). A recognition test was used for measuring factual knowledge; a comprehension test for measuring conceptual understanding; and a transfer test for measuring higher-level understanding on the taxonomy levels of evaluation, synthesis, and creation.

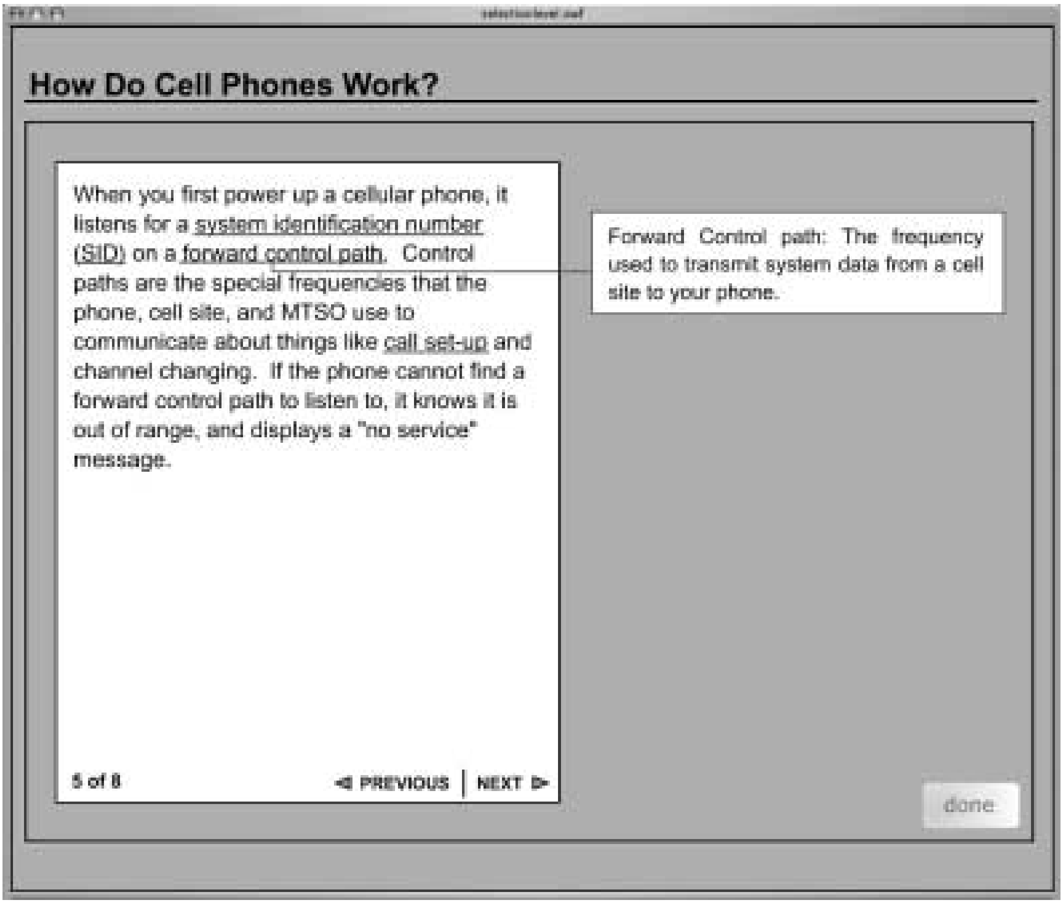
The participant questionnaire solicited demographic information. The survey of prior knowledge asked questions regarding participants' existing knowledge about cell phone technology. The online instructional text consisted of 650 words that explained the operation of cell phone systems, and was developed using Flash (Macromedia, 2003) and ColdFusion (Macromedia, 2002). Three types of text annotations were developed for the reading text. Each level of annotation was designed to support one of the three cognitive processes described in Mayer's (2001) cognitive theory of multimedia learning:

1. Twenty-one selection-level annotations were designed to support the selection of relevant information, and annotated individual words or concepts, such as *forward control path,* by providing a definition of the term without contextual information (see Appendix A for examples).
2. Nine organization-level annotations were designed to support the connection of words into ideas, and annotated phrases or ideas, such as the function and organization of an individual cell, by providing a brief explanation of the idea in the specific context.
3. Six integration-level annotations were designed to support the construction of connections between the different ideas and prior knowledge, and annotated entire paragraphs by showing the links of the ideas in the paragraph.

The total amount of text in each type of annotation was roughly the same. That is, the total number of words in all of the selection-level annotations was roughly equal to the total number of words in all of the organization-level annotations, which were roughly equal to the total number of words in all of the integration-level annotations.

The software program was designed so that the text appeared on the left side of the screen and the annotations appeared on the right (see Figure 1). The text and annotations were positioned close to each other to avoid split attention, consistent with previous studies on text annotations (Chandler & Sweller, 1992; Plass et al., 2003). Participants accessed each annotation by clicking on an underlined portion of the text (for selection- and organization-level annota-

Figure 1 ☐ Screen shot of the reading text with selection-level annotation selected for the word *cell.*



tions; integration-level annotations were marked by a link to the right of the text). Participants were able to access only one annotation at a time, but were able to refer back and forth between their selected annotation and the original text as much as they wished. They could access the annotations at any time during their reading of an individual page of text and they could read each annotation as many times as they wished. Participants were asked to access all available annotations and those who accessed fewer than 75% of the annotations were excluded from the analysis.

Seven versions of the program were developed for each group, and contained the same 650-word text on four pages; the only difference in the versions was in the number and type of available annotations. Version NA (corresponding to the control group) contained only the 650-word text.; Version S had additional annotations for individual words. Version O had annotations for groups of words (idea units). Version I had annotations for entire paragraphs. Version SO had annotations both for words and idea units. Version SI had annotations for both words and paragraphs. Version OI had annotations for both idea units and entire paragraphs.

Learning outcomes were measured with the three different tests mentioned above. Similar tests have been successfully used in previous research (Mayer, 2001; Mayer & Moreno, 1998; Moreno & Mayer, 2000). The recognition test consisted of 21 multiple-choice items that assessed participants' knowledge of cell phone systems. The comprehension test consisted of 6 cued-recall items designed to measure participants' understanding (comprehension) of each of the subsystems involved in a cell phone system. Participants were asked to write down as many answers as they could think of for each of the questions. For instance, 1 question asked participants to "Write down everything you can remember about cells and cell sites." The transfer test was designed to measure how complete and accurate participants' mental models of how cell phone systems work were. This test consisted of 5 items that asked participants to use their knowledge in new and creative ways that were not covered in the reading text. For instance, 1 question asked participants to imagine that they were riding a train across the country, making cell phone calls as the connection occasionally went dead. They were asked to write down all of the reasons they could think of why this might occur.

The paper-based materials consisted of a test of verbal ability—the Opposites Test—from the ETS Kit of factor-referenced cognitive tests (Ekstrom, French, Harman, & Dermen, 1976).

Procedure

All procedures were performed in a computer lab with small groups of participants. After completing a form consenting to participate in this research, participants were randomly assigned to one of the seven groups. After first completing a form requesting demographic information and then the prior knowledge test online, they reviewed a Web page with instructions on how to use the reading text and annotations, and were allowed to ask any question they might have regarding the procedure. They were then given as much time as they needed to read the text about cell phones and look up any annotations. After completing the instruction, participants completed the three Web-based learning outcome tests (recognition, comprehension, and transfer). Next, participants were given 4 min for each of the two parts of the Opposites Test (Ekstrom, et al., 1976). Finally, participants were thanked and debriefed.

Scoring

The scoring of all materials was done by one or more scorers who were not aware of the participants' treatment conditions. The recognition test was scored by assigning 1 point to each of the correct responses on the multiple-choice test, yielding a maximum score of 21. The comprehension test was scored by assigning 1 point to each correct idea unit from the text learners had written down, which yielded a maximum score of 27. For the transfer test, participants received 1 point for each acceptable correct (nontrivial) answer across all of the transfer test questions; the maximum score in this test was 15.

RESULTS

Research Question 1—Effects of Different Types of Annotations

Will different types of annotations result in differential levels of learning outcomes comprehending a scientific text? Based on Mayer's (2001) cognitive theory of multimedia learning, we expected that annotations could be designed with a specific function to support text comprehension on a higher or lower level.

To address this question, we compared the results of the recognition test, comprehension test, and transfer test for the four treatment conditions. We computed a 4×2 multivariate analysis of variance (MANOVA) with group as between-subject factor with four levels (NA, S, O, I), verbal ability as between-subject variable with two levels (low, high), created with a median split, and with the three outcome measures (recognition, comprehension, transfer) as dependent variables. Table 1 shows the descriptives for the four treatments and three outcome measures. Results show multivariate effects for group (Wilks's $\Lambda = .482$, $F = 4.35$, $df = 9$, $p < .0001$, $\eta_p{}^2 = .216$) and for verbal ability (Wilks's $\Lambda = .818$, $F = 3.40$, $df = 3$, $p < .05$, $\eta_p{}^2 = .182$). Univariate analyses for each of the three outcome measures showed that group differences were statistically significant for the recognition test ($F(3, 56) = 11.16$, $MSE = 1{,}714.21$, $\eta_p{}^2 =$

Table 1 ☐ Mean percent correct and standard deviation on recognition test, comprehension test, and transfer test for groups receiving no annotations (NA-Controls) or one type of annotation (S, O, or I).

Test & Group	*N*	*Mean*	*Standard Deviation*
Recognition Test			
NA	14	32.99	13.78
S	14	59.86	11.62
O	15	50.48	12.83
I	14	43.88	15.31
Comprehension Test			
NA	14	11.94	8.97
S	14	19.69	10.75
O	15	20.19	9.03
I	14	13.16	8.76
Transfer Test			
NA	14	19.16	12.12
S	14	33.12	16.67
O	15	36.36	11.40
I	14	29.87	10.35

NA = no annotations; S = selection; O = organization; I = integration

.411, $p < .0001$), the comprehension test ($F(3, 56) = 2.89$, $MSE = 225.25$, $\eta_p{}^2 = .153$, $p < .05$), and the transfer test ($F(3, 56) = 5.12$, $MSE = 779.87$, $\eta_p{}^2 = .242$, $p < .005$).

Post hoc comparisons of recognition test scores revealed that participants recognized significantly more words from the text when they received any of the three types of annotations than participants in the control group with no annotations ($p < .05$) (see Table 1). In addition, only learners who received selection-level annotations outperformed learners who received integration-level annotations on this test. However, the difference between learners receiving selection-level annotations and those receiving organization-level annotations did not reach statistical significance.

For the comprehension test, post hoc comparisons of test scores revealed that learners who received selection-level and organization-level annotations recalled more ideas from the text than learners who received integration-level annotations or no annotations ($p < .05$) (see Table 1). In addition, only learners who received organization-level annotations recognized more words than learners who received integration-level annotations; the difference between learners receiving organization-level annotations and learners receiving selection-level annotations did not reach statistical significance.

Finally, for the transfer test, post hoc comparisons of test scores revealed that, compared to the control group, students generated significantly more creative solutions when they received any of the three types of annotations. The learners receiving selection-, organization- or integration-level annotations did not differ from one another (see Table 1).

These results indicate that there is indeed a differential impact of the different types of annotations on the different learning outcome measures. Consistent with our predictions, selection-level and organization-level annotations improved learning in all three measures, and integration-level annotations resulted in increased transfer test scores, but also helped with the recognition of words from the text.

Research Question 2—Effects of Multiple Annotations

Will the availability of two different types of annotations result in increased or decreased learning as compared to one type of annotation? We expected that learners receiving two types of annotations would learn less than learners receiving only one type of annotation because the low prior knowledge of the learners and the difficulty of the material would induce a higher cognitive load.

In order to test this hypothesis, we compared the results of the recognition text, comprehension test, and transfer test for treatments where one type of annotation was available with those for treatments that included combinations of two types of annotations (selection + organizing, selection + integrating, organizing + integrating). For that purpose, we coded three dummy factors, one for each type of annotation. Each of these factors (SEL, ORG, INT) had two levels, representing the absence (0) or presence (1) of the respective selection-level, organization-level, or integration-level annotation. We computed a $2 \times 2 \times 2$ MANOVA with these three factors (SEL, ORG, INT) as between-subject factors and our three outcome measures (recognition, comprehension, transfer) as dependent variables. The MANOVA model specified for this analysis included main effects for all factors and interaction effects for all pairs of annotation types: SEL + ORG, SEL + INT, ORG + INT.

The multivariate results show main effects for SEL, the availability of selection-level annotations (Wilks's $\Lambda = .744$, $F = 10.07$, $df = 3$, $p < .01$, $\eta_p{}^2 = .256$), and interaction effects for SEL × ORG, combinations of selection- and organization-level annotations (Wilks's $\Lambda = .877$, $F = 4.12$, $df = 3$, $p < .01$, $\eta_p{}^2 = .123$), and ORG × INT, combinations of organization- and integration-level annotations (Wilks's $\Lambda = .879$, $F = 4.03$, $df = 3$, $p < .01$, $\eta_p{}^2 = .121$).

Univariate analyses for each of the three outcome measures showed that with the inclusion of groups receiving more than one type of annotation per treatment condition, the results regarding the impact of different types of annotation on different outcome measures (that we found previously when we analyzed results for groups receiving only one type of annotation per treatment condition) had mostly disappeared: There was a main effect only for the availability of selection-level annotation on the recognition test ($F(1, 90) = 10.23$, $MSE = 1{,}889.87$, $\eta_p{}^2 = .102$, $p < .01$), indicating that these annotations led to improved performance on this test.

To answer our research question, which was concerned with the question of whether the absence or presence of one type of annotation modified the effect of another type of annotation on learning, we focused on the interaction effects found in the univariate analyses. Because these interactions involve combinations of two of the dummy factors we used to describe the treatment conditions, that is, the absence or presence of specific types of annotations, an interaction effect is an indication for differences in learning outcomes as a result of how having no type of annotation, either one type of annotation, or both types of annotations available. Tables 2, 3, and 4 show the descriptives for the three outcome measures in the different treatment combinations.

For the recognition test, there were statistically significant interaction effects of selection-level and organization-level annotations ($F(1, 90) = 7.61$, $MSE = 1{,}405.91$, $\eta_p{}^2 = .078$, $p < .01$) and of organization- and integration-level annotations ($F(1, 90) = 4.34$, $MSE = 819.97$, $\eta_p{}^2 = .038$, $p < .01$). Pairwise comparisons revealed that for selection- and organization-level annotations, receiving either type of annotation resulted in learners recognizing significantly more words from the text compared to receiving no annotations ($p < .01$). Learners receiving two types of annotations did not recognize more words than those receiving no annotation. The difference between learners receiving two types of annotations and learners receiving either one type of annotation only, did not reach statistical significance, however. Similar results were found for organization- and integration-level annotations: Learners receiving organization-level annotations recognized more words from the text than learners not receiving any annotations ($p < .05$), for integration-level annotations that difference was only a trend ($p = .072$); learners receiving both types of annotations did not differ in their test performance from learners with no annotations, see Tables 2, 3, and 4. Again, the difference between learners receiving both types of annotations and those receiving only one type did not reach statistical significance.

For the comprehension test, there was a significant interaction of selection-level and organization-level annotations ($F(1, 90) = 9.10$, $MSE = 822.88$, $\eta_p{}^2 = .092$, $p < .01$). Pairwise comparisons revealed, similar to the pattern obtained for the recognition test, that learners who received organization-level annotations recognized significantly more words from the text than those who received no annotations ($p < .05$); for selection-level annotations this effect was marginally significant ($p = .60$). Learners who received both types of annotations performed significantly worse on the comprehension test than learners

Table 2 ☐ Mean percent correct and standard error on recognition test, comprehension test, and transfer test for groups receiving no annotations (controls), one type of annotation (S or O), and two types of annotations (SO).

Dependent Variable	*Group*	*N*	*Mean*	*Standard Error*	*95% Confidence Interval Lower Bound*	*95% Confidence Interval Upper Bound*
Recognition Test	NA	14	39.19	2.42	34.37	44.00
	S	14	61.04	2.30	56.46	65.62
	O	15	51.86	2.46	46.96	56.75
	SO	10	52.49	5.83	40.90	64.08
Comprehension Test	NA	14	13.30	1.65	10.02	16.59
	S	14	17.44	1.57	14.31	20.57
	O	15	19.33	1.68	15.98	22.67
	SO	10	6.83	3.98	–1.09	14.74
Transfer Test	NA	14	25.10	2.22	20.67	29.52
	S	14	31.94	2.12	27.74	36.15
	O	15	33.78	2.26	29.28	38.27
	SO	10	15.16	5.35	4.51	25.80

Estimated marginal means
NA = no annotations; S = selection; O = organization; SO = selection and organization

Table 3 ☐ Mean percent correct and standard error on recognition test, comprehension test, and transfer test for groups receiving no annotations (controls), one type of annotation (S or I), and two types of annotations (SI).

Dependent Variable	*Group*	*N*	*Mean*	*Standard Error*	*95% Confidence Interval Lower Bound*	*95% Confidence Interval Upper Bound*
Recognition Test	NA	14	43.10	2.42	38.29	47.92
	S	14	58.68	2.67	53.38	63.99
	I	14	47.94	2.46	43.04	52.84
	SI	17	54.85	5.34	44.23	65.47
Comprehension Test	NA	14	16.62	1.65	13.33	19.91
	S	14	15.62	1.82	12.00	19.25
	I	14	16.01	1.68	12.66	19.35
	SI	17	8.64	3.65	1.39	15.90
Transfer Test	NA	14	28.69	2.22	24.27	33.12
	S	14	29.97	2.45	25.10	34.84
	I	14	30.18	2.26	25.68	34.68
	SI	17	17.13	4.90	7.37	26.88

Estimated marginal means
NA = no annotations; S = selection; I = integration; SI = selection and integration

Table 4 □ Mean percent correct and standard error on recognition test, comprehension test, and transfer test for groups receiving no annotations (controls), One type of annotation (O or I), and two types of annotations (OI).

Dependent Variable	*Group*	*N*	*Mean*	*Standard Error*	*95% Confidence Interval*	
					Lower Bound	*Upper Bound*
Recognition Test	NA	14	47.12	2.42	42.31	51.94
	O	15	54.66	2.67	49.36	59.97
	I	14	53.11	2.30	48.53	57.69
	OI	14	49.68	5.55	38.65	60.72
Comprehension Test	NA	14	16.46	1.65	13.17	19.75
	O	15	15.79	1.82	12.16	19.41
	I	14	14.29	1.57	11.16	17.41
	OI	14	10.36	3.79	2.83	17.90
Transfer Test	NA	14	26.99	2.22	22.57	31.41
	O	15	31.68	2.45	26.80	36.55
	I	14	30.05	2.12	25.84	34.26
	OI	14	17.26	5.10	7.12	27.39

Estimated marginal means
NA = no annotations; O = organization; I = integration; OI = organization and integration

who received either one type of annotation ($p < .05$), see Tables 2, 3, and 4.

For the transfer test, there were interactions of all three combinations of annotations (selection- or organization-level: $F(1, 90) = 10.41$, $MSE = 1{,}671.80$, $\eta_p{}^2 = .104$, $p < .01$; selection- and integration-level: $F(1, 90) = 4.40$, $MSE = 705.88$, $\eta_p{}^2 = .047$, $p < .05$; organization- and integration-level: $F(1, 90) = 8.55$, $MSE = 1{,}372.86$, $\eta_p{}^2 = .087$, $p < .01$). Supplemental pairwise comparisons showed that for each of these interactions, learners receiving two types of annotations generated significantly fewer creative solutions than those receiving one type of annotation ($p < .05$), see Tables 2, 3, and 4.

Though the differences did not reach significance for all learning outcome measures and combinations of annotations, multiple annotations resulted in lower mean scores on several outcome measures, lending support to our hypothesis that additional textual annotations led to a higher amount of extraneous cognitive load that resulted in lower knowledge acquisition.

Research Question 3—Moderating Effects of Verbal Ability

Does verbal ability have a moderating effect on the load experienced in learning with multiple annotations? We expected that low-verbal ability learners would experience higher load in learning with multiple annotations than high-verbal ability learners, especially under those conditions where load was already high.

Because the number of participants in this study was not high enough to answer this research question, we were able to conduct the comparison of learners with different levels of verbal ability only as an exploratory analysis. Verbal ability was entered into the design used for Research Question 2 as a between-subject factor with two levels by using a median split of the verbal ability scores, generating a high-verbal and a low-verbal ability group. This factor was then added to the MANOVA model computed in the previous section. The resulting exploratory $2 \times 2 \times 2 \times 2$ MANOVA showed, as one would expect, the same patterns of multivariate main effects and interaction effects as the MANOVA model in the previous section. There were a multivariate main effect for SEL, the availability of selection-level annotations (Wilks's $\Lambda = .729$, $F = 10.05$, $df = 3$, $p < .01$, $\eta_p^2 = .271$), and multivariate interaction effects for SEL × ORG, combinations of selection- and organization-level annotations (Wilks's $\Lambda = .824$, $F = 5.77$, $df = 3$, $p < .01$, $\eta_p^2 = .176$), and ORG × INT, combinations of organization- and integration-level annotations (Wilks's $\Lambda = .886$, $F = 3.46$, $df = 3$, $p < .05$, $\eta_p^2 = .114$). In addition, however, we found a main effect for verbal ability (Wilks's $\Lambda = .787$, $F = 7.29$, $df = 3$, $p < .0001$, $\eta_p^2 = .213$) and three-way interactions of verbal ability and the availability of organization- and integration-level annotations (Wilks's $\Lambda = .910$, $F = 2.65$, $df = 3$, $p < .01$, $\eta_p^2 = .090$).

ANOVAs showed main effects for verbal ability for all outcome measures (recognition: $F(1, 83) = 10.4$, $MSE = 2{,}198.56$, $\eta_p^2 = .14$, $p < .001$, comprehension: $F(1, 83) = 18.07$, $MSE = 1{,}355.49$, $\eta_p^2 = .179$, $p < .001$; transfer: $F(1, 83) = 14.14$, $MSE = 1{,}917.34$, $\eta_p^2 = .146$, $p < .001$). They also revealed, for the transfer test, significant three-way interactions of treatments involving combinations of organization- and integration-level annotations with verbal ability ($F(1, 83) = 7.50$, $MSE = 1{,}017.20$, $\eta_p^2 = .083$, $p < .01$) as well as of treatments involving combinations of selection- and organization-level interactions with verbal ability ($F(1, 83) = 4.19$, $MSE = 567.88$, $\eta_p^2 = .048$, $p < .05$).

Supplemental pairwise comparisons revealed that for both of these three-way interactions, low-verbal-ability learners scored lower than high-verbal-ability learners in the groups that received two types of annotations ($p < .05$). For treatment groups receiving one type of annotation, there were no differences in scores for high- versus low-verbal-ability learners.

DISCUSSION

In this research, we addressed the question of whether or not different types of annotations, designed to support different levels of text processing, would result in different learning outcomes. We found that, consistent with our hypothesis and with Mayer's (2001) cognitive theory of multimedia learning, annotations designed to support the process of selecting information were the most effective annotations for the recognition of words from the text and their definitions, followed by annotations designed to support the process of organizing information. A similar pattern emerged for the comprehension test, in which participants who received annotations designed to support the process of organizing information and annotations designed to support the process of selecting information achieved higher performance on the comprehension test than did those who did not. In the comprehension tests, annotations designed to support the process of integrating information did not lead to higher test performance than receiving no annotations. In the recognition and transfer tests, however, these annotations led to significantly higher scores than those received by the control group. In the transfer test, selection- and organization-level annotations also led to performance that was better than that of the control group. This was expected, as support in the earlier stages of processing should result in improvements in higher-level learning. Conversely, however, support in the later stages of processing, as expected, did not result in improved lower-level learning. This result pattern supports the underlying hypothesis of generative processes in text comprehension (Wittrock, 1990) and cognitive theory of multimedia learning (Mayer, 2001).

In addition to the main effects for one type of annotation, we also examined the effect of providing learners with two types of annotations, and the resulting cognitive load implications. Here we found that, with the exception of one main effect for selection-level annotations in the recognition tests, all of the main effects described above had disappeared as a result of having two types of annotations available. Instead, we found several interaction effects, each showing that combinations of the annotations that, as single annotations, had significantly improved learning, now resulted in lower performance. The fact that this effect was particularly consistent for the transfer test, a measure of higher-level processing of information, suggests that this is a cognitive load effect: The requirements of processing two types of annotations generated extraneous load that exceeded learners' cognitive capacity. While this increase in extraneous load had a smaller impact on rote learning and comprehension, it hampered learners' deeper processing of the materials.

Our third, exploratory, research question investigated the impact of verbal ability on the effectiveness of the annotations. We found that learners with higher verbal ability showed, in general, a higher performance than learners with lower verbal ability. In addition, we found that the lower performance on the test involving higher-level thinking found for treatment conditions with two types of annotations was further exacerbated for learners with a low verbal ability, whose performance in treatments with multiple annotations was worse than that of learners with a high verbal ability.

The conclusions of this study have to be tempered with some caveats. First, cognitive load was not measured directly, so we cannot know what the actual levels of cognitive load were. However, this is the case in most studies that investigate the impact of cognitive load effects on knowledge acquisition (Brünken et al., 2003). Second, although the number of words for each type of annotation was similar, the number of words varied across conditions where either one or two types of annotations were available. The total number of words in the conditions receiving two types of annotation was greater than the total number of words in the single annotation conditions, which had more words than the no annotation controls. The observed cognitive load effect in the multiple annotation conditions might be caused by the additional amount of textual information that had to be processed. However, if this were indeed the case, then one would expect all combinations of any two types of annotations to induce the same level of high cognitive load, which was not the case in our data. Third, our sample size was relatively low for the analyses involving verbal ability as a factor. Although group differences were statistically significant, it would have been desirable to have a larger sample size for these analyses. Fourth, we investigated the impact of textual annotations only on cognitive load and knowledge acquisition. Future research needs to be conducted to examine to what degree the observed effect pattern can be generalized to other annotations using other presentation modes, such as pictures, videos, or animation.

This research offers theoretical as well as practical implications. On the theoretical side, it supports the notion of assigning a particular function to annotations used in reading text (Levin et al., 1987), and redefines these functions by specifying the cognitive processes of multimedia learning that each type of annotation supports (Mayer, 2001). This research further shows that there is no case in which multiple annotations led to higher performance than single annotations, indicating that, for learners with low prior knowledge, the cognitive load of processing multiple annotations exceeds their available cognitive capacity. Future research is needed to study in more detail the impact of verbal ability on cognitive load. On the practical side, our research provides instructional designers with guidance regarding the design of annotations for scientific texts, allowing them to consider the cognitive function of these annotations, and the level of cognitive load they will generate for learners with different levels of verbal ability. □

Erik Wallen [ewallen@salud.unm.edu] is with the Health Sciences Center at the University of New Mexico.

Jan L. Plass [jan.plass@nyu.edu] is with the Educational Communication and Technology Program, School of Education, New York University.

Roland Brünken is with the Department of Psychology at Georg August University of Göttingen, Germany.

Correspondence concerning this article should be addressed to Erik Wallen, University of New Mexico, Health Sciences Center, Department of Internal Medicine, Albuquerque, NM 87131, or to Jan L. Plass, New York University, The Steinhardt School of Education, program in Edcuational Communication and Technology, 239 Green St., Suite 300, New York, NY 10003.

REFERENCES

Anderson, L. W., & Krathwohl, D. R. (Eds.). (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. New

York: Longman.

Brünken, R., Plass, J. L., & Leutner, D. (2003). Direct measurement of cognitive load in multimedia learning. *Educational Psychologist, 38*, 5361.

Brünken, R., Plass, J. L., & Leutner, D. (2004). Assessment of cognitive load in multimedia learning with dual task methodology: Auditory load and modality effects. *Instructional Science*, *32*, 115–132.

Carney, R. N., & Levin, J. R. (2002). Pictorial illustrations *still* improve students' learning from text. *Educational Psychology Review, 14*, 526.

Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and Instruction, 8*, 293332.

Chandler, P., & Sweller, J. (1992). The split attention effect as a factor in the design of instruction. *British Journal of Educational Psychology, 62*, 233246.

Chun, D. M., & Plass, J. L. (1996). Effects of multimedia annotations on vocabulary acquisition. *The Modern Language Journal, 80*, 183198.

Chun, D. M., & Plass, J. L. (1997). Research on text comprehension with multimedia. *Language Learning and Technology [Online serial], 1*(1), 60–81.

Ekstrom, R. B., French, J. W., Harman, H. H., & Dermen, D. (1976). *Kit of factor-referenced cogitive tests*. Princeton, NJ: Educational Testing Service.

Jones, L. C., & Plass, J. L. (2002). Supporting listening comprehension and vocabulary acquisition with multimedia annotations. *The Modern Language Journal, 86*, 546–561.

Levin, J. R., Anglin, G. J., & Carney, R. R. (1987). On empirically validating functions of pictures in prose. In D. M. Willows & H. A. Houghton (Eds.), *The psychology of illustration: Vol. I. Basic research* (pp. 51–85). New York: Springer.

Macromedia. (2002). ColdFusion Server MX [Computer Software]. San Francisco: Author.

Macromedia. (2003). Flash 5 [Computer Software]. San Francisco: Author.

Mayer, R. E. (2001). *Multimedia learning*. New York: Cambridge University Press.

Mayer, R. E., & Moreno, R. (1998). Split-attention effect in multimedia learning: Evidence for dual processing systems in working memory. *Journal of Educational Psychology, 90*, 312–320.

Moreno, R., & Mayer, R. E. (2000). A coherence effect in multimedia learning: The case for minimizing irrelevant sounds in the design of multimedia instructional messages. *Journal of Educational Psychology, 92*, 117–125.

Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive load theory and instructional design: Recent developments. *Educational Psychologist, 38*, 1–4.

Paivio, A. (1986). *Mental representations: A dual-coding approach.* New York: Oxford University Press.

Plass, J. L., Chun, D. M., Mayer, R. E., & Leutner, D. (1998). Supporting visual and verbal learning preferences in a second-language multimedia learning environment. *Journal of Educational Psychology, 90*, 25–36.

Plass, J. L., Chun, D. M., Mayer, R. E., & Leutner, D. (2003). Cognitive load in reading a foreign language text with multimedia aids and the influence of verbal and spatial abilities. *Computers in Human Behavior, 19*, 221–243.

Sweller, J. (1999). *Instructional Design in Technical Areas*. Camberwell, Victoria, Australia: Australian Council for Educational Research.

Sweller, J., Chandler, P., Tierner, P., & Cooper, M. (1990). Cognitive load in the structuring of technical material. *Journal of Experimental Psychology: General, 119*, 176–192.

Sweller, J., Van Merriënboer, J. \ J. G., & Paas, F. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10*, 251–295.

Wittrock, M. C. (1990). Generative Processes of Comprehension. *Educational Psychologist*, *24*, 345–376.

Appendix overleaf.

Appendix A ☐ Examples of text annotations.

Selection-level annotation: *Forward control path: The frequency used to transmit system data from a cell site to your phone.*

Organization-level annotation: *The forward control path is the data channel used to send system information from cell site to cell phones. To identify themselves, cell sites continuously broadcast their system identification number over a forward control path, enabling the cell phone to determine its own location in the service area.*

Integration-level annotation: *A cell phone needs to know whether it is in an area where it can make and receive calls. To do this, it looks for active cell sites by scanning forward control paths in the same way that you might scan the radio dial for a radio station. If you find a station, you can judge from the quality of the sound how good the connection with the station is. If the cell phone finds an active forward control path, it knows that it is in range and can make and receive calls. Most cell phones have a signal strength display that indicates how good the reception of the signal on the forward control path is.*