

**GENERATING TOPIC HEADINGS DURING READING
OF SCREEN-BASED TEXT FACILITATES LEARNING
OF STRUCTURAL KNOWLEDGE AND IMPAIRS
LEARNING OF LOWER-LEVEL KNOWLEDGE**

ROY B. CLARIANA

Penn State University

ANTHONY W. MARKER

Boise State University

ABSTRACT

This investigation considers the effects of learner-generated headings on memory. Participants ($N = 63$) completed a computer-based lesson with or without learner-generated text topic headings. Posttests included a cued recall test of factual knowledge and a sorting task measure of structural knowledge. A significant disordinal interaction was observed with the no headings group performing better on the cued recall test and worse on the structural knowledge sorting task. In addition, a measure of structural knowledge using a list-wise rating task was field trialed. The two structural knowledge tasks were analyzed using a Pathfinder network scoring technique that measured both linear and nonlinear text subtopic organization. The sorting and rating tasks were differentially sensitive to participants' linear and nonlinear knowledge organization. Unexpectedly, the sorting and rating posttest tasks improved cued recall posttest scores. The implications of the results for measuring structural knowledge are considered and suggestions for further studies are presented.

Generative learning approaches are strategies that encourage learners to actively consider the relationships among lesson information elements and between lesson information and personal knowledge (Jonassen, 1988; Jonassen & Wang, 1993; Wittrock, 1992). For example, generating headings, formulating new questions,

forming direct inferences, and drawing concept maps, require learners to consider multiple information elements in a text, thus encouraging the development of the organizational and structural relationships between information elements (Grabowski, 1996; Ritchie & Volkl, 2000). Generative strategies may positively affect traditional posttest measures but due to their nature, are also likely to affect more holistic knowledge structure, such as *structural knowledge* (Jonassen, Beissner, & Yacci, 1993; Mason-Mason & Tessmer, 2000; McKeague & Di Vesta, 1996).

Structural knowledge, also known as cognitive structure, refers to the organization of information in memory. Jonassen et al. (1993) suggests that well developed structural knowledge is a critical intermediate for applying declarative knowledge thus emphasizing why further research on generative learning approaches is important.

Author-Provided Headings as Signals

In the headings literature, the term “topic” refers to the content of a section of text and “heading” is the descriptive label for the topic. In this investigation, the term “subtopic” will refer to ideas, propositions, micro-propositions, and other information in the text that is subordinate to a topic. Viewed as an outline of the text passage, topic headings are the major points in the outline while subtopics are the points under each topic.

Typically, headings inserted in text by the author are reading signals, an approach that is usually not considered to be a generative strategy (Wilhite, 1988) though McKeague and Di Vesta (1996) have described headings as an organization aid that highlights the potential structure of a text. Two leading explanations of author-provided headings as signals for text processing are the *selection* hypothesis and the *implementation* hypothesis. The selection hypothesis (Loman & Mayer, 1983; Lorch & Lorch, 1995; Mayer, Dyck, & Cook, 1984; Meyer, Brandt, & Bluth, 1980; Meyer & Rice, 1982, 1989) maintains that when text is unsignaled, readers primarily use a “linear” strategy, encoding the text information as an ordered list of ideas. When text-topic signals such as headings are present, most mature readers adopt a “structure” strategy (Meyer, 1984), encoding the text as *clustered* sets of meaningfully related ideas. In contrast, the implementation hypothesis (Lorch & Lorch, 1985, 1995) maintains that mature readers tend to use a structure strategy whether signals are present in the text or not. Both hypotheses assume that author-provided text-topic headings in a text influence which text ideas (i.e., topics and subtopics) are included in memory after reading and also how these ideas are structured (i.e., linear or clustered).

Most previous research on author-provided headings (see Sanchez, Lorch, & Lorch, 2001) has shown that readers remember about the same number of subtopics in signaled and unsignaled conditions. However, readers in the signaled condition recall a broader selection of topics (referred to as topic access) and topic

mention more closely follows the linear order of occurrence of the topics in the lesson text (referred to as topic organization and also as topic order). If author-generated headings have this impact on reader's retention of the text, what then is the likely affect of learner-generated headings?

Learner-Generated Headings

In order to generate a meaningful topic heading, the learner must review the section of text while creating a heading for that section; essentially, the learner must engage in a *structure strategy approach*. As a result of this close reading, subtopics occurring in the same text section should be more tightly organized or clustered together in memory.

Previous investigations suggest that learners who generate headings score *lower* than no-headings control groups on factual recall tests and possibly other lower-order outcomes, but score *higher* on inference and comprehension tests (Dee-Lucas & DiVesta, 1980; Jonassen, Hartley, & Trueman, 1985; McKeague & Di Vesta, 1996; Wittrock & Kelly, 1984). These interesting negative effects of generating headings on lower-level learning outcomes are similar to Reyna and Brainerd (1998) findings that verbatim and gist traces (fuzzy trace theory) compete for activation at retrieval, and gist traces intrude at posttest (Clariana & Koul, 2006; Seamon, Luo, Schwartz, Jones, Lee, & Jones, 2002). On the other hand, generating headings while reading positively affects higher-order outcomes, and in our view, this is due to the generative effects of headings on readers' structural knowledge. How do you measure structural knowledge?

Measuring Structural Knowledge

Typically, the effects of headings on memory are assessed by asking participants to recall all the topics and subtopics from the lesson text that they can remember; and in some studies, by asking participants to summarize the lesson text which is then converted to a list of topics and subtopics by the researcher. This list of subtopics and topics is then analyzed to determine if the order of first mention of the subtopics matches the topic order of the lesson text. A topic is considered to be recalled if the participant recalls at least one of its subtopics at posttest. For example, if a participant recalls three subtopics in a row at posttest that are all subsumed under the same topic, this counts as one topic recall. Previous analysis methods using Spearman correlation of topic order recall provide a measure of linear topic recall, but are not a very rich measure of structural knowledge.

Based on Jonassen's et al. (1993) approaches for eliciting and measuring structural knowledge, two approaches were used in this investigation, a previously validated semantic map sorting task (Taricani & Clariana, 2006) and a new approach employing a semantic judgment abbreviated list-wise rating task

(Clariana & Wallace, in press). The tasks were designed to measure both linear and nonlinear (i.e., clustering) subtopic organization.

Most headings research has involved print-based rather than screen-based text. Schultz and Spyridakis (2004) and Spyridakis, Schultz, and Bartell (2005) compared the affects of headings in screen-based compared to print-based text. They reported that under certain conditions, headings in screen-based text (in this case, a long scrolling text with various frequencies of headings) have very different effects on reading comprehension relative to print-based text (i.e., specifically, a negative effect on comprehension with high-frequency headers). Thus previous print-based text investigations of the effects of both author-provided and learner-generated headings should not be automatically generalized to all kinds of screen-based text.

This investigation focuses on the affects of a generative strategy, reader-generated text topic headings while reading screen-based text. We asked, does generating headings while reading improve memory of verbatim information in the text, relative to reading without headings? And does generating headings while reading induce structural memory that more closely reflects the topic and subtopic structure of the text, relative to reading without headings? We anticipate a disordinal interaction, with the headings group scoring lower on the verbatim posttest and higher on the structural knowledge posttest.

Posttest measures include a sorting task measure of structural knowledge and a cued recall declarative knowledge (fill-in-the-blank) measure of memory of information found in the lesson text. In addition, an untried structural knowledge rating task was compared to the previously established sorting task to determine its efficacy and validity for measuring structural knowledge.

METHOD

Participants

Participants in this investigation were adult professional students ($N = 63$) enrolled in several different instructional design graduate courses. Participants were school teachers or else were employed full time in training-related positions in the corporate, non-profit, or higher education sectors. Participants were in their mid-thirties and most were female (65%).

Participation was voluntary. In some cases, students could choose to participate in the investigation as one of several alternative research-related course projects. In other cases, the instructional materials provided a review of instructional design vocabulary for the introductory course, and students completed the materials as a review. The confidentiality of participants' posttest performance was maintained by using an anonymous numerical identification system.

Materials

The computer software randomly assigned participants to one of two lesson treatments, Headings or No Headings. Both the Headings and the No Headings lesson treatments began with a flowchart diagram of the eight components of the Dick and Carey (1996) model of instructional design, which is the model used in the introductory course. Next, students read an 801 word text summary of this topic. The text was organized into seven topic sections, with each section displayed on a separate screen. The seven topic sections followed the order of the Dick and Carey model and included: (a) Needs Assessment (with 4 subtopics), (b) Domains of Instructional Outcomes (5 subtopics), (c) Instructional Analysis (4 subtopics), (d) Learner and Context Analysis (5 subtopics), (e) Objectives and Posttests (4 subtopics), (f) Instructional Strategy (5 subtopics), and (g) Formative Evaluation (3 subtopics).

Spyridakis et al. (2005) categorized heading frequency as low frequency (approximately 300 words per heading), medium frequency (approximately 200 words per heading), and high frequency (approximately 100 words per heading). Previous headings-as-signals investigations have used text of different lengths with quite variable numbers of headings, but a typical study involves a text of about 2000 words with about 20 headings. This ratio of 133 words per topic heading (i.e., 2000/15) would be categorized as high frequency, and practically speaking, amounts to about one heading per paragraph. The topic heading frequency in the present investigation is 801 words divided by 7 headings equals 114 words per topic heading, which would also be categorized as high frequency. Practically, this amounted to one heading per screen of text.

To compel interactivity with the text during the lesson, the most important concept terms (the subtopics) for each section were listed in the margin on the left side of the screen, and the text passage contained blank spaces. Participants in both the Headings and No Headings treatments were instructed to “Click and drag the term on the left to the correct blank in the text.” Instructional feedback consisted of a “try until correct” approach. When correct, the term locked into the blank, but when incorrect, the term flew across the screen to its original starting place on the left side of the screen. Thus, participants must correctly place all terms into the text passage section before moving on to the next section.

The two treatments were identical, except that in the Headings treatment, after completing each section, participants were directed to generate a heading for that section. Generating their own unique heading for a lesson text section should encourage the participant to associate the content and subtopics in that section together in a personally meaningful way (Lorch & Lorch, 1996). The Headings treatment instructions stated, “Now, type in a Heading for this section, and then press Enter,” and a red arrow pointed to a blank text entry box located at the top of the section. Participants in the Headings treatment could not advance to the next section until they typed something into the headings text entry box.

Procedure

Participants logged in to the software program in any of the campus computer labs or classes. The software randomly assigned the participant to the Headings or else the No Headings treatment. On completing the lesson treatment, participants were then randomly assigned by the software to receive the rating or else the sorting structural knowledge posttest task first, and then the remaining rating or sorting task second. Finally, participants completed the fill-in-the-blank posttest.

Cued Recall Posttest

The cued recall declarative knowledge posttest consisted of 30 fill-in-the-blank instructional design terminology questions taken from a previous study by Clariana and Lee (2001). The questions were intended to measure lower-order, verbatim declarative factual knowledge of the lesson text content. Each definition was presented on the screen one at a time in random order, and the participants were asked to type in the correct term. The cued recall posttest consisted of two 15-item subtests, referred to as *Subtest 1* and *Subtest 2*. *Subtest 1* consisted of 15-items taken from lesson text sections B, D, and F; while *Subtest 2* consisted of 15 questions taken from lesson text sections A, C, E, and G. The 15 items in *Subtest 1* aligned with the 15 subtopics that were included in the sorting and rating posttest tasks (see the next section), while the 15 items in *Subtest 2* were *not* aligned with any subtopics included in the sorting and rating posttest tasks. Cronbach reliability for the full 30-item declarative knowledge posttest is $\alpha = 0.83$. Cronbach reliability for *Subtest 1* is $\alpha = .85$ and for *Subtest 2* is $\alpha = .74$.

Sorting Task Posttest

The sorting task posttest is a computer-delivered semantic mapping task that is scored using a Pathfinder associative network technique, as described by Taricani and Clariana (2006). The computer program developed for this investigation for collecting sorting task (and rating task) raw data is the Knowledge Unit Mapper (*KU-Mapper*, 2003). Previous pilot testing showed that sorting 30 subtopics on one screen is complex and potentially overwhelming for the user. Thus, the sorting task screen displayed only 15 of the 30 subtopics in the lesson text. The 15 subtopics to be sorted were taken from lesson text sections B, D, and F (the second, fourth, and sixth lesson text sections). Directions on the screen stated, “Drag related terms closer together and unrelated terms further apart. When done, click CONTINUE” (see Figure 1).

The *KU-Mapper* software was used to convert the sorting task raw data into proximity array files. The values in the array are the *distances* between terms measured in screen pixels. An array containing 15 terms requires 105 elements in order to represent every list-wise relationship between the 15 terms (e.g., $(15^2 - 15)/2 = 105$).

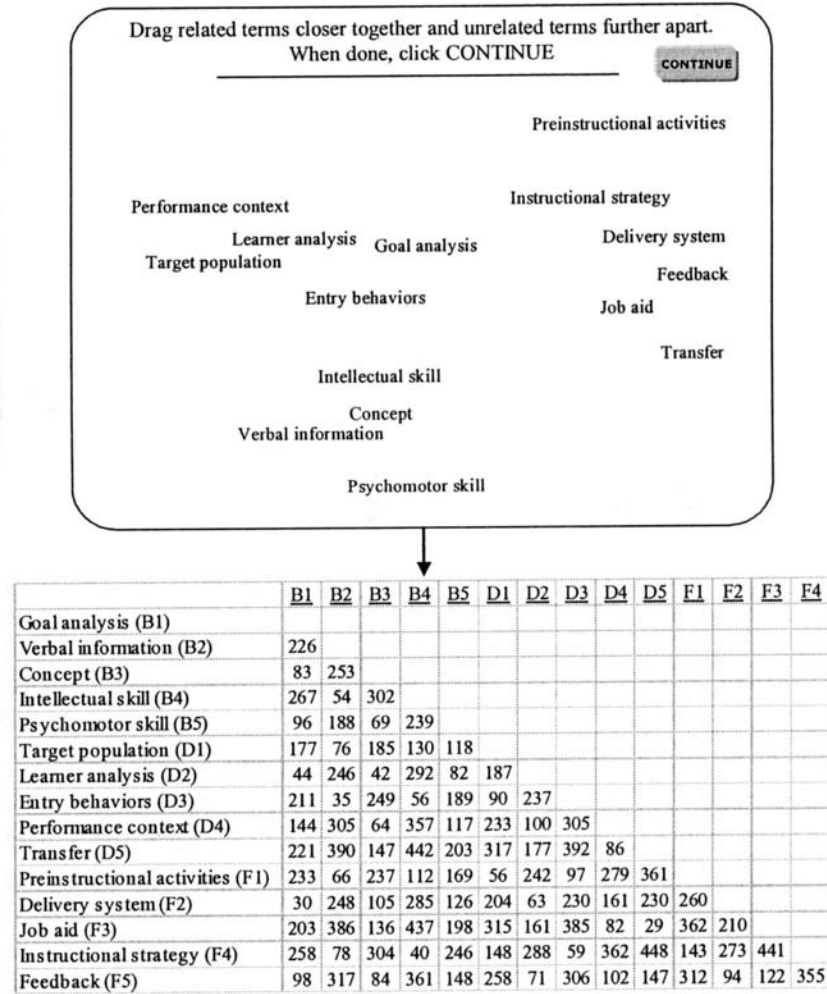


Figure 1. A participant's sorting task posttest (above) and the resulting distance array (below). B, D, and F designate the lesson sections and 1 through 5 indicate serial order within that lesson text section.

The distance array files were analyzed by software called Knowledge Network and Orientation Tool for the Personal Computer, version 4.3 (*KNOT*, 1998). First, *KNOT* software is used to reduce the raw distance array data into a Pathfinder network representation (*PFNet*). *PFNets* are visual (and mathematical) representations of the most salient aspects of the distance raw data. The most restricted *PFNets* contain about $n - 1$ links (where n is the number of terms used). Then *KNOT* software was used to score the resulting *PFNets*. The scoring technique is described later.

Experimental Rating Task Posttest

In addition, a computer-based abbreviated list-wise rating task was specifically designed for this investigation as an alternative way to measure relationships between subtopics (*KU-Mapper*, 2003). The task is also based on the Pathfinder multiple pair-wise rating approach, but is somewhat more like a tree construction test (see Jonassen et al., 1993, p. 53) except that, rather than drawing links between most related terms, participants select the most related terms from a list. The list-wise rating task posttest used the same 15 subtopics as the semantic map sorting task posttest described above so that data from the two approaches for measuring structural knowledge could be directly compared.

The directions on the first screen of the list-wise rating task stated, “Your task is to ‘click-on’ the concept on the right side that is most related to the target concept shown on the left side. We are most interested in your impression of ‘overall relatedness’, so base your selections on your first impression of relatedness. You will make 15 such comparisons.” Next, one randomly selected subtopic was displayed on the left-hand side of the screen, a statement in the middle of the screen said “is most related to,” and a list of subtopics was displayed on the right-hand side of the screen in alphabetical order (see Figure 2). Directions above the list stated, “click one of the terms below.” After selecting the most related subtopic from the list, another subtopic was randomly displayed on the left-hand side and so on until all 15 subtopics had been used exactly once.

The list-wise rating task software was designed so that “self-self” comparisons are not possible. For example, if the subtopic “job aid” is displayed on the left-hand side, then it is not displayed on the right-hand side. The software also dynamically alters the right-hand list to exclude previously associated subtopics. For example, if a participant has already selected “learner analysis” from the list as being most related to “target audience,” then later when “learner analysis” is displayed on the left-hand side, “target audience” will not appear as a choice on the right-hand side. In that case, rather than 14 subtopics appearing on the list on the right-hand side, only 13 subtopics would be displayed. This approach may or may not be optimal, but was adopted in this case in order to increase the amount of link raw data in order to allow the algorithm more flexibility when deriving the Pathfinder network path.

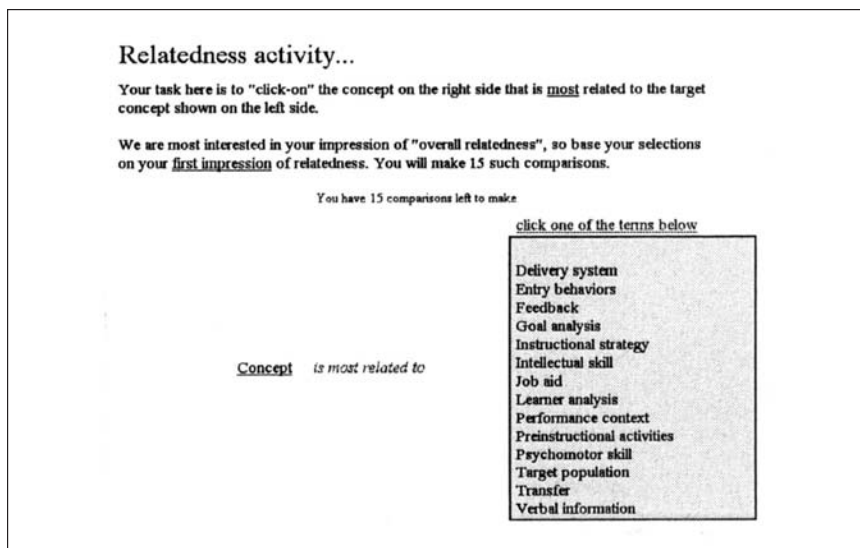


Figure 2. The list-wise rating task posttest screen layout.

The list-wise rating task software generates proximity array files that can be analyzed by *KNOT* software. Like the distance arrays for the semantic map sorting task, the rating task arrays contain 105 elements to represent every list-wise proximity relationship between the terms, however the rating task array contains only "1s" (related) and "0s" (not related).

Scoring the Sorting and Rating Task Posttests

Linear and Nonlinear Referents

Participants' scores are established by counting the number of links in their *PFNets* that agree with two referent *PFNets*, a linear referent and a nonlinear (i.e., cluster) referent. Thus for scoring purposes, a linear referent was established by creating an array file containing "1"s to designate every possible valid link between *contiguous* lesson text subtopics and a nonlinear referent was established by creating an array file containing "1"s to designate every possible valid link between lesson text subtopics within a lesson text section but that are *not* in serial order (see Figure 3).

For example, the linear referent contains a link between the B1 and B2 subtopics (e.g., Goal analysis and Verbal information), between the B2 and B3 terms subtopics, and so on, but not between the B1 and B3 subtopics. Since there are three lesson text sections, referred to here as "B", "D" and "F," with five terms in each section, then 12 links are required (i.e., 4 + 4 + 4) to account for the lesson

	Linear Referent					Nonlinear Referent				
	B1	B2	B3	B4	B5	B1	B2	B3	B4	B5
Goal analysis (B1)	—					—				
Verbal information (B2)	1	—				0	—			
Concept (B3)	0	1	—			1	0	—		
Intellectual skill (B4)	0	0	1	—		1	1	0	—	
Psychomotor skill (B5)	0	0	0	1	—	1	1	1	0	—

Figure 3. Linear and nonlinear referent data arrays for lesson Section B.

text subtopic serial order (see the left panel of Figure 3). Similarly, the nonlinear referent contains a link between the B1 and B3 subtopics but not between the B1 and B2 subtopics. Since there are three lesson text sections with five subtopics in each section, then 18 links are required (i.e., $6 + 6 + 6$) to account for all possible nonlinear effects of the lesson treatments on subtopic knowledge organization (see the right panel of Figure 3).

Converting Array Files into Posttest Scores

KNOT software was used to convert the participants' sorting and rating array files into *PFNets*. The software requires several input parameters that are used to describe the raw data and the desired analysis approach. In this investigation the following parameters were used, Minkowski's r was set equal to infinity and q was set equal to 14 (i.e., number of subtopics minus 1). In addition, for the sorting task raw data, the data were defined as dissimilarity data, the maximum value was set to 1000, and the minimum value was set to 0. For the rating task raw data, the data were defined as similarity data, the maximum value was set to 1, and the minimum value was set to 0.1 (i.e., with this unusual data set, the minimum must be set to a value slightly greater than 0, otherwise spurious links between section subtopics will be established).

After all of the *PFNets* were formed, *KNOT* software was used to calculate the number of links shared in common for each participant's *PFNets* compared to the linear and the nonlinear referents. Links in common are the scores obtained for the sorting and rating posttests.

RESULTS

The results section consists of several analyses including: (a) Repeated-measures analysis of the declarative and structural knowledge posttests to address the two questions posed in this investigation, (b) *Post hoc* analysis of *Subtest 1*

due to unexpected findings in the first repeated-measures analysis, (c) Correlation analysis of the two measures of structural knowledge to determine their reliability and construct-related equivalency, and (d) Repeated-measures analysis of the sorting and rating structural knowledge tasks to consider the efficacy and validity of the new abbreviated rating task scores. The means and standard deviations for the posttest are shown in Table 1.

Comparison of the Declarative and Structural Knowledge Posttests

To place the cued recall (declarative knowledge) and sorting task (structural knowledge) data on the same scale of measurement, the raw scores for each measure were converted to percent scores by dividing by the total possible (maximum 30 and 14 respectively). The means and standard deviations for the No Headings control group are $M = 0.63$ and $SD = 0.20$ for cued recall, and $M = 0.45$ and $SD = 0.11$ for the sorting task; and for the Headings experimental group, $M = 0.55$ and $SD = 0.24$ for cued recall, and $M = 0.50$ and $SD = 0.16$ for the sorting task. Pearson's correlation between the cued recall and sorting task scores was significant but weak ($r = 0.34, p = .006$).

The cued recall and sorting task posttest data were analyzed by a 2 (Treatment: Headings vs. No Headings) \times 2 (Posttest: cued recall and sorting task) mixed ANOVA. The first is a between-subjects factor and the second is the within-subjects factor. The Treatment main effect was not significant, $F(1, 61) = 0.220$, $MSE = 0.045$, $p = .94$. The Posttest repeated measure was significant, $F(1, 61) = 18.874$, $MSE = 0.022$, $p < .001$, showing that the mean cued recall test score ($M = 0.59$) was greater than the mean sorting task score ($M = 0.47$). Finally, the anticipated disordinal interaction of Treatment and Posttest factors was significant, $F(1, 61) = 5.119$, $MSE = 0.022$, $p = .027$ (see Figure 4).

Table 1. Means and Standard Deviations^a for the Linear and Nonlinear Sorting and Rating Tasks and for the Cued Recall Subtest 1 and Subtest 2

Treatment	Cued Recall		Sorting task		Rating task	
	Subtest 1	Subtest 2	Linear	Nonlinear	Linear	Nonlinear
No Headings $n = 32$	10.3 (3.6)	8.6 (3.1)	2.9 (1.1)	3.4 (1.3)	4.0 (1.6)	3.6 (1.5)
Headings $n = 31$	9.0 (4.0)	7.5 (3.3)	2.9 (1.3)	4.0 (1.6)	3.8 (1.6)	3.2 (1.7)

^aStandard deviation shown in parentheses.

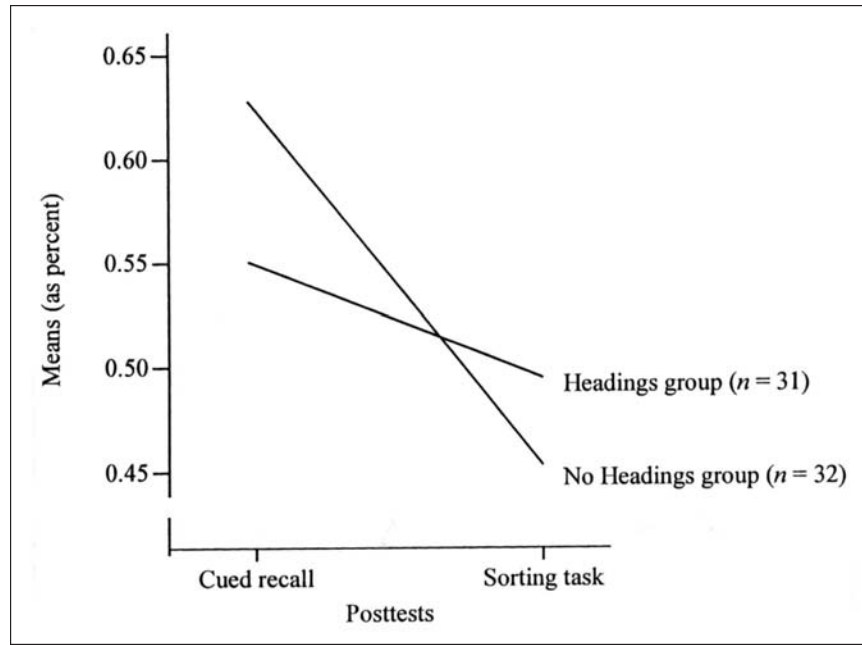


Figure 4. The significant disordinal interaction of headings treatment and posttest.

Asking learners to generate topic headings tended to suppress cued recall scores, but tended to improve structural knowledge scores. Simple main effects follow-up show that the Headings and No Headings groups were *not* significantly different on the cued recall test ($p = .052$) or the sorting task ($p = .063$). Though not significant, the No Headings group performed a little better than the Headings group on the cued recall posttest while the Headings group structural knowledge was a bit more like the lesson text structure than the No Headings group.

Post hoc Analysis of Subtest 1

While inspecting the cued recall posttest scores, we noticed that *Subtest 1* scores were substantially greater than *Subtest 2* scores. In order to interpret this apparent difference, it is necessary to know whether the two subtests mean scores are expected to be equivalent. The posttest item difficulties for the 30 fill-in-the-blank questions used in the present investigation were reported earlier in a study by Clariana and Lee (2001). The treatments and participants from these two investigations are comparable in that the test questions in both studies are

identically worded, the participants were drawn from the same population, and the treatments were somewhat similar screen-based lessons.

For *Subtest 2* (the 15 fill-in-the-blank items that are *not* aligned with the sorting and rating tasks), the Clariana and Lee (2001) average item difficulty was 0.54, while the average item difficulty in the present investigation is 0.53, virtually the same. However, for *Subtest 1* (the 15 fill-in-the-blank items that are aligned with the intervening sorting and rating posttest tasks), the Clariana and Lee (2001) average item difficulty was 0.42 while the average item difficulty in the present study is 0.63, an effect size gain of 0.94, which with this sample size, is significant at the $p < .001$. This suggests that the sorting and rating intervening posttest tasks facilitated fill-in-the-blank posttest performance only for *Subtest 1*, those items that were aligned with subtopics that were actually sorted and rated. There was *not* a generalized affect for sorting and rating for *Subtest 2*.

Correlation Analysis of the Sorting and Rating Tasks

The concurrent criterion-related validity of the sorting and rating tasks measures of knowledge structure has theoretical and pragmatic importance. The sorting and rating subtests (i.e., linear and nonlinear scores) were compared to *Subtest 1* (i.e., the fill-in-the-blank questions that aligned with lesson text sections B, D, and F and with the sorting and rating posttest content). Two of the four structural knowledge measures were significantly related to the cued recall *Subtest 1* ($p < .05$), the rating task linear subtest ($r = 0.50$) and the sorting task nonlinear subtest ($r = 0.28$). The rating task nonlinear subtest ($r = 0.24$) and the sorting task linear subtest ($r = 0.17$) were not significantly related to the cued recall subtest.

The rating and sorting tasks combined score (i.e., linear plus non-linear) are significantly related ($r = 0.42$). However, the sorting and rating linear and nonlinear subtests are not significantly related to each other, except for the sorting and rating tasks linear subtest measure ($r = 0.26$). These low and non correlations indicate that the sorting and rating subtests are not measuring the same thing. This is difficult to interpret, but apparently the two tasks elicit linear and nonlinear memory associations differently.

Analysis of the Sorting and Rating Structural Knowledge Posttest Tasks

The sorting and rating tasks linear and nonlinear data were analyzed by a 2 (Treatment: Headings vs. No Headings) \times 2 (Task: sorting and rating) \times 2 (Structure: Linear and Nonlinear) mixed ANOVA, the first is a between subjects factor and the second and third are within-subjects factors. The treatment main effect was not significant, $F(1, 61) = 0.007$, $MSE = 3.106$, $p = .94$, indicating that overall, there was no difference between the Headings and No Headings combined linear and nonlinear map and rating task means (see Table 1).

The posttest Task repeated measure was significant, $F(1, 61) = 6.918$, $MSE = 1.200$, $p = .011$, with the rating-task mean significantly greater than the sorting-task mean ($M = 14.6$ compared to 13.2). The Treatment by Task interaction was significant, $F(1, 61) = 4.044$, $MSE = 1.200$, $p = .049$ (see Figure 5).

Scheffé analysis was used to follow up this interaction since all pairs of means were examined. This analysis indicates that the No Headings group rating task mean was significantly greater than its sorting task mean while the Headings group obtained nearly identical sorting and rating task means. None of the other pair-wise comparisons were significant.

The Structure repeated measure was not significant $F(1, 61) = 0.470$, $MSE = 2.460$, $p = .49$ nor was the Treatment by Structure interaction, $F(1, 61) = 0.210$, $MSE = 2.460$, $p = .65$. However, the interaction of the posttest Task and Structure repeated measures was significant, $F(1, 61) = 13.471$, $MSE = 2.051$, $p = .001$ (see Figure 6).

Follow-up Scheffé analysis indicates that the sorting-task linear mean was significantly less than both the rating-task linear mean and also the sorting-task nonlinear mean. Finally, the three way interaction of Treatment, Task, and

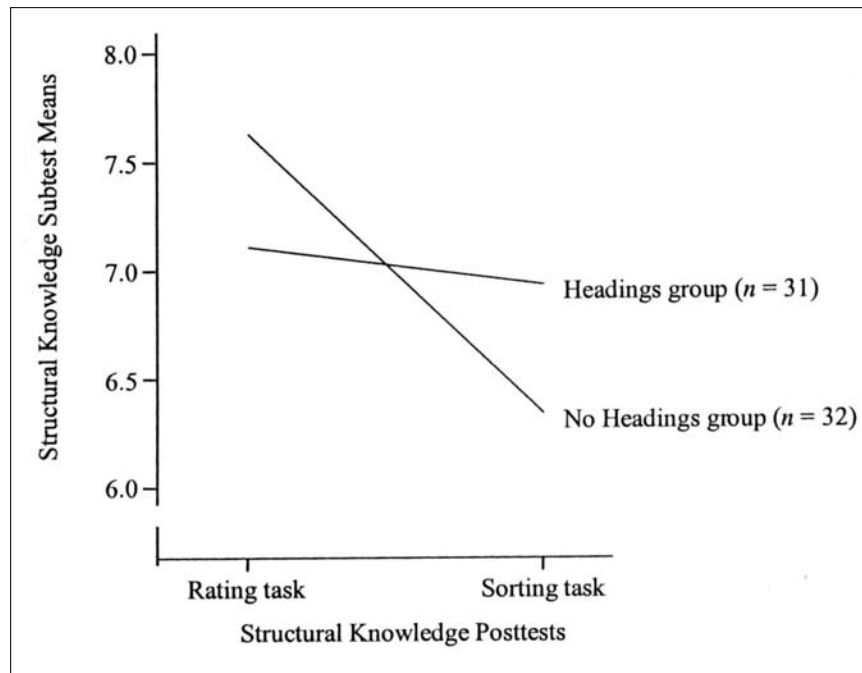


Figure 5. The significant interaction of headings treatment with sorting and rating tasks.

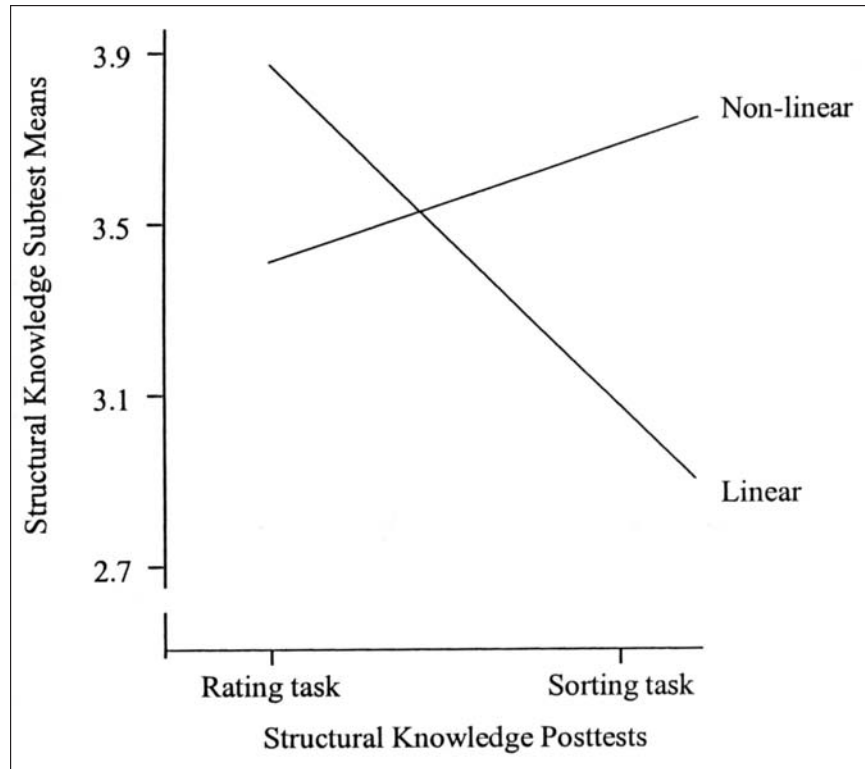


Figure 6. The significant interaction of Task and Structure.

Structure was not significant, $F(1, 61) = 0.973$, $MSE = 2.051$, $p = .33$. Apparently, either the sorting task does not elicit recall of the text subtopic linear order, or else the sorting task is not a good measure of text subtopic linear order relative to the rating task, though the sorting task does seem to be a relatively better measure of nonlinear subtopic order.

DISCUSSION

This investigation considered the effects of learner-generated headings on two measures of structural knowledge and on a fill-in-the-blank posttest. Results indicate that generating headings while reading did not improve memory of verbatim information in the text or substantially alter structural knowledge relative to reading without headings. However, although not significant, the No Headings group performed a little better than the Headings group on the lower-order cued recall posttest ($p = 0.052$) while the Headings group structural knowledge

(i.e., sorting task combined linear and non-linear measure) was more like the lesson text structure ($p = 0.063$) than the No Headings group (see Figure 4). These results are consistent with and help explain previous investigations that have reported that learners who generate headings score *lower* than no-headings control groups on lower-order outcomes but score *higher* on inference and comprehension tests (Dee-Lucas & DiVesta, 1980; Jonassen et al., 1985; Wittrock & Kelly, 1984). The better structured knowledge of the Headings group (i.e., more like the author's text schema) should allow the learners to more flexibly use that knowledge (Jonassen & Wang, 1993) which should affect the reader's ability to form inferences and comprehend the lesson text, but this apparently comes at the expense of text details. Perhaps generating headings induces a stronger fuzzy (e.g., gist) trace that intrudes on verbatim memory traces of text details at posttest (Clariana & Koul, 2006; Reyna & Brainerd, 1998).

Unexpectedly, the intervening sorting and rating structural knowledge posttests facilitated improved cued recall (fill-in-the-blank) posttest scores. Future research should consider whether the intervening posttest effect observed here is due to the sorting task, the rating task, or the combination of the two. For example, the linear subtest of the list-wise rating task had the highest correlation with the cued recall posttest ($r = .50$), thus the list-wise rating approach may be primarily responsible for this intervening posttest effect. Further research should consider which aspect of either task may have had this powerful influence on verbatim cued recall; or alternately, this effect may be due just to a general practice effect.

The sorting and rating tasks appear to elicit and/or measure different aspects of structural knowledge and in addition the two tasks are differentially sensitive to participants' linear and nonlinear knowledge organization. The sorting task appears to be relatively more sensitive to nonlinear lesson text memory associations relative to the rating task. Thus the sorting task may not be as good as the rating task at measuring linear structural knowledge but would be better than the rating task at measuring nonlinear structural knowledge. As the participants' knowledge structure becomes less linear but more "clustered" (i.e., the Headings group), then the rating and sorting task posttest measures become more related.

The brief text used in this investigation (801 words) may limit the generalizability of these findings. Future research should replicate this investigation but use longer text passages (i.e., 5000 words) and add a higher-order posttest measure, for example, a comprehension, applications, or problem-solving test.

A critical caveat in this investigation is that the participants were familiar with the text topic. McKeague and Di Vesta (1996) have reported that learners who are novice or unfamiliar with the text content and are asked to generate text topic headings perform significantly lower on both verbatim and text organization posttest measures compared to similar participants in an author-provided treatment. Thus instructional designers must consider the learners' entry level of content knowledge (i.e., learners must be familiar with the content) before selecting learner-generated headings as an appropriate instructional activity.

Practically speaking, generating headings while reading appears to strengthen (relatively) the nonlinear, whole section relationships that are likely established during the generation of headings but also depress verbatim declarative knowledge (see Figure 4). Instructional design of screen-based text for higher-order learning outcomes should probably include learner-generated headings but designers should *not* use learner-generated headings for lower-order outcomes. Also, including the rating and/or sorting tasks (the posttest measures of structural knowledge in this investigation) as *instructional tasks* should strongly improve cued recall memory of lesson information, possibly without interfering with higher-order learning outcomes.

In conclusion, more research is needed to determine the intermediate affects of generative learning strategies on structural knowledge and on various lower- and higher-order outcomes. Perhaps the terms “fuzzy trace” and “structural knowledge” refer to the same underlying mental phenomenon. Can fuzzy trace theory be used to inform generative learning strategies? Studies using magnetic resonance imaging (MRI) of the effects of headings while reading suggest that the middle temporal regions of the right hemisphere may be especially important for integrative processes needed to achieve global coherence during reading (St. George, Kutas, Martinez, & Sereno, 1999). Expanding both the theoretical underpinnings and ways of measuring both processing effects and learning outcomes should lead to better understanding of the effects of generative learning strategies.

REFERENCES

- Clariana, R. B., & Koul, R. (2006). The effects of different forms of feedback on fuzzy and verbatim memory of science principles. *British Journal of Educational Psychology*, 76(2), 259-270.
- Clariana, R. B., & Lee, D. (2001). The effects of recognition and recall study tasks with feedback in a computer-based vocabulary lesson. *Educational Technology Research and Development*, 49(3), 23-36.
- Clariana, R. B., & Wallace, P. E. (in press). A comparison of pair-wise, list-wise, and clustering approaches for eliciting structural knowledge. *International Journal of Instructional Media*. Downloaded December 6, 2006 from http://www.personal.psu.edu/rbc4/ku_mapper.doc
- Dee-Lucas, D. & DiVesta, F. F. (1980). Learner-generated organizational aids: Effects on learning from text. *Journal of Educational Psychology*, 72(3), 304-311.
- Dick, W., & Carey, L. (1996). *The systematic design of instruction* (4th ed.). New York: HarperCollins.
- Grabowski, B. L. (1996). Generative learning: Past, present, future. In D. H. Jonassen (Ed.), *Handbook for educational communications and technology*, pp. 897-918. New York: Simon and Schuster Macmillan.
- Jonassen, D. H. (1988). *Instructional designs for microcomputer courseware*. Hillsdale, NJ: Lawrence Erlbaum Associates.

- Jonassen, D. H., Hartley, J., & Trueman, M. (1985, April). *The effects of learner generated versus experimenter-provided headings on immediate and delayed recall and comprehension*. Chicago: American Educational Research Association (ERIC ED 254 567).
- Jonassen, D. H., & Wang, S. (1993). Acquiring structural knowledge from semantically structured hypertext. *Journal of Computer-Based Instruction*, 20(1), 1-8.
- Jonassen, D. H., Beissner, K., & Yacci, M. (1993). *Structural knowledge: Techniques for representing, conveying, and acquiring structural knowledge*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- KNOT. (1998). *Personal computer knowledge network organization tool, version 4.3*. Downloaded April 17, 2003 from <http://interlinkinc.net/>
- KU-Mapper. (2003). *Knowledge unit mapper*. Downloaded December 1, 2006 from www.personal.psu.edu/rbc4
- Loman, N. L., & Mayer, R. E. (1983). Signaling techniques that increase the understandability of expository prose. *Journal of Educational Psychology*, 75, 402-412.
- Lorch, R. F., & Lorch, E. P. (1985). Topic structure representation and text recall. *Journal of Educational Psychology*, 77, 137-148.
- Lorch, R. F., & Lorch, E. P. (1995). Effects of organizational signals on text processing strategies. *Journal of Educational Psychology*, 87, 537-544.
- Lorch, R. F., & Lorch, E. P. (1996). Effects of headings on text recall and summarization. *Contemporary Educational Psychology*, 21, 261-278.
- Mason-Mason, S. D., & Tessmer, M. (2000). Expert systems as a mindtool to facilitate mental model learning. *Educational Technology Research and Development*, 48(4), 43-62.
- Mayer, R. E., Dyck, J. L., & Cook, L. K. (1984). Techniques that help readers build mental models from scientific text: Definitions pretraining and signaling. *Journal of Educational Psychology*, 76, 1089-1105.
- McKeague, C. A., & Di Vesta, F. J. (1996). Strategy orientations, learner activity, and learning outcomes: implication for instructional support of learning. *Educational Technology Research and Development*, 44(2), 29-42.
- Meyer, B. J. F. (1984). Text dimensions and cognitive processing. In H. Mandl, N. L. Stein, & T. Trabasso (Eds.), *Learning and comprehension of text*. Hillsdale, NJ: Erlbaum.
- Meyer, B. J. F., Brandt, D. M., & Bluth, G. J. (1980). Use of top-level structure in text: Key for reading comprehension of ninth-grade students. *Reading Research Quarterly*, 16, 72-103.
- Meyer, B. J. F., & Rice, E. (1982). The interaction of reader strategies and the organization of text. *Text*, 2, 155-192.
- Meyer, B. J. F., & Rice, E. (1989). Prose processing in adulthood: The text, the reader and In L. W. Poon, D. C. Rubin, & B. A. Wilson (Eds.), *Everyday cognition in adult and later life* (pp. 157-194). New York: Cambridge University Press.
- Reyna, V. F., & Brainerd, C. J. (1998). Fuzzy-trace theory and false memory: New frontiers. *Journal of Experimental Child Psychology*, 71, 194-209.
- Ritchie, D., & Volkl, C. (2000). Effectiveness of two generative learning strategies in the science classroom. *School Science and Mathematics*, 100(2), 83-89.
- Sanchez, R. P., Lorch, E. P., & Lorch, R. F. (2001). Effects of headings on text processing strategies. *Contemporary Educational Psychology*, 26, 418-428.

- Schultz, L. D., & Spyridakis, J. H. (2004). The effect of heading frequency on comprehension of online information: a study of two populations. *Technical Communication, 51*(4), 504-516.
- Seamon, J. G., Luo, C. R., Schwartz, M. A., Jones, K. J., Lee, D. M., & Jones, S. J. (2002). Repetition can have similar or different effects on accurate and false recognition. *Journal of Memory and Language, 46*, 323-340.
- Spyridakis, J. H., Schultz, L. D., & Bartell, A. L. (2005). Heading frequency and comprehension: Studies of print versus online media. *Theory, Research, Education, and Training, nv*, 138-142. Online version downloaded September 16, 2006 from <http://www.stc.org/ConfProceed/2005/PDFs/0032.pdf>
- St. George, M., Kutas, M., Martinez, A., & Sereno, M. I. (1999). Semantic integration in reading: Engagement of the right hemisphere during discourse processing. *Brain, 122*, 1317-1325.
- Taricani, E. M., & Clariana, R. B. (2006). A technique for automatically scoring open-ended concept maps. *Educational Technology Research and Development, 54*, 61-78.
- Wilhite, S. C. (1988). Reading for a multiple-choice test: Headings as schema activators. *Journal of Reading Behavior, 20*(3), 215-228.
- Wittrock, M. C. (1992). Generative learning processes of the brain. *Educational Psychologist, 27*(4), 531-41.
- Wittrock, M. C., & Kelly, R. (1984). *Teaching reading comprehension to adults in basic skills courses*. Final Report, Project No. MDA 903-82-C-0169). University of California, Los Angeles.

Direct reprint requests to:

Roy B. Clariana
Associate Professor of Education
The Pennsylvania State University
1510 South Quebec Way #42
Denver, CO 80231
e-mail: RClariana@psu.edu

Copyright of Journal of Educational Computing Research is the property of Baywood Publishing Company, Inc. and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.

Copyright of Journal of Educational Computing Research is the property of Baywood Publishing Company, Inc. and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.