Instr Sci (2010) 38:679–705
DOI 10.1007/s11251-009-9095-8

What makes a matrix so effective? An empirical test of the relative benefits of signaling, extraction, and localization

Douglas F. Kauffman · Kenneth A. Kiewra

Received: 27 October 2008 / Accepted: 30 March 2009 / Published online: 22 April 2009

Abstract What type of display helps students learn the most and why? This study investigated how displays differing in terms of signaling, extraction, and localization impact learning. In Experiment 1, 72 students were assigned randomly to one cell of a 4 × 2 design. Students studied a standard text, a text with key ideas extracted, an outline that localized ideas topically, and a matrix that localized ideas topically and categorically. One version of the displays signaled the displays' organization and one version did not. The matrix display proved best for facilitating fact and relationship learning because of its ability to localize related information within topics and categories. Simply signaling or extracting text ideas was not helpful. Experiment 2 demonstrated that not all matrices are created equal because they can vary in terms of how information is localized. About 54 students were assigned randomly to one cell of a 2 × 2 design that varied localization of matrix topics and categories. Students studied matrices high or low in topical organization and high or low in categorical organization. Results confirmed that a high, natural ordering of matrix topics is necessary to highlight relationships and bolster relationship and fact learning.

Keywords Text processing · Matrix organizer · Studying · Study materials · Graphic organizers

D. F. Kauffman (✉)
Department of Educational Psychology, University of Nebraska, 222 Teachers College Hall, Lincoln, NE 68588-0345, USA
e-mail: dkauffman2@unl.edu

K. A. Kiewra
Department of Educational Psychology, University of Nebraska, 240 Teachers College Hall, Lincoln, NE 68588-0345, USA
e-mail: kkiewra1@unl.edu

Introduction

Suppose you are a science teacher asking students to read the simple text about wildcats found in Fig. 1. What might students learn?

Students might learn discrete facts about wildcats such as *tigers live in jungles* and *lions live in groups*, but might not learn implicit relationships among facts, such as *cats that live in the jungle are solitary, whereas cats that live on the plains live in groups*. To find this relationship, a reader must extract relevant facts from multiple text locations, organize them, and then identify the relationship. Unfortunately, readers are unlikely to perform these operations even with text as simple and brief as Fig. 1 (e.g., Jairam and Kiewra in press; Kauffman 2004).

Now suppose the identical wildcat information is presented in the matrix display shown in Table 1. Studying this matrix, students can still learn discrete facts by reading down wildcat columns. It is easy to see, for example, that *tigers roar* and that *bobcats live six years.* But the matrix's two-dimensional structure allows learners to do more than learn discrete facts. The matrix can also be read horizontally to compare wildcats along common categories such as genus and call. By reading across multiple wildcat categories, several relationships are easily discerned, such as: (a) *cats from the genus Panthera roar, whereas cats from other genuses hiss and purr;* (b) *cats that live in the jungle are solitary, whereas cats that live on the plains reside in groups;* and (c) *heavier cats live longer than lighter weight cats.* Given the matrix display, readers need not manipulate text to discern relationships; the underlying relationships are apparent almost at once.

Wildcats

The tiger is classified into the genus Panthera. Its most common call is its ferocious roar. The tiger's social behavior is solitary and its habitat is the jungle. The tiger has a life span of 25 years and can weigh up to 450 pounds.

The lion is a member of the genus Panthera and its most common call is its mighty roar. Lions roam the plains habitat during their 25-year life span. The adult lion weighs about 400 pounds. It is a social animal that lives in a group.

The cheetah belongs to the genus Acinonyx. It lives in social groups and its habitat is the plains. The cheetah has a life span of eight years. Its most common calls are the hiss and purr, and its maximum weight is 125 pounds.

The bobcat's life span in its jungle habitat is six years. Its social behavior is defined by its solitary nature. The bobcat belongs to the genus Lynx, its primary calls are the hiss and purr, and its maximum weight is 30 pounds.

Fig. 1 Simple wildcat text

Table 1 Simple wildcat matrix

	Tiger	Lion	Cheetah	Bobcat
Genus	Panthera	Panthera	Acinonyx	Lynx
Call	Roar	Roar	Hiss and purr	Hiss and purr
Weight	450	400	125	30
Life span	25	25	8	6
Habitat	Jungle	Plains	Plains	Jungle
Social behavior	Solitary	Groups	Groups	Solitary

The present study investigated the learning potential of text and matrix displays but also examined how other displays—extracted text, outlines, and displays with or without signaling—differentially affect student learning. Although past research has favored matrix displays over text and outline displays (e.g., Kiewra et al. 1999; McCrudden et al. 2004; Robinson et al. 1998), it has not compared these displays with others or fully explained why one display is better than another. Below is a brief description of each of the displays investigated in the present study and its inclusion rationale.

Display characteristics

Standard text is linear and in paragraph form thereby forcing readers to follow a single processing path: left-to-right and top-to-bottom (Kiewra et al. 1999). This linear processing path makes it difficult for readers to locate and understand relationships among facts dispersed throughout the text. Consequently, readers learn discrete facts rather than existing relationships (Kauffman 2004). Standard text was used in the present study to determine what students learn without the benefit of text aids.

One compensation for text's linear presentation is a *signaled text* that cues the reader's attention to the text's underlying structure (Lorch 1989). The present study examined whether text signals in the form of boldface type, italics, and underlining, when used in text or other displays, help the reader recognize the organizational structure and existing relationships. Figure 2 presents a signaled version of the wildcat text presented earlier. Note that the topics (i.e., cats' names) are in boldface type; the categories (e.g., genus and call) are in italics; and the corresponding facts are underlined. These cues signal text structure and should help readers to link facts (450 pounds) with corresponding topics (tiger) and categories (weight) and to discern the text's overriding structure: four wildcats each described with regard to six common categories. Discerning the text's structure might also help readers recognize relationships among wildcats such as *the tiger is the heaviest cat.* In the present study, the independent effects of signaling were examined by comparing the learning potential of displays with or without signaling.

Although the signaled text highlights the text's structure, it fails to extract important information and organize it so that relationships are more visible. Related facts, such as the common calls of the cheetah and bobcat, remain embedded within the larger text and are separated by intervening and excess information. Although the signaled text should draw the reader's attention to the highlighted facts about each wildcat, its comprehensiveness and block-like structure might still limit relationship learning.

Wildcats

The **tiger** is classified into the *genus* Panthera. Its most common *call* is its ferocious roar. The tiger's *social behavior* is solitary and its *habitat* is the jungle. The tiger has a *life span* of 25 years and can *weigh* up to 450 pounds.

The **lion** is a member of the *genus* Panthera and its most common *call* is its mighty roar. Lions roam the plains *habitat* during their 25 year *life span*. The adult lion *weighs* about 400 pounds. It is a *social* animal that lives in a group.

The **cheetah** belongs to the *genus* Acinonyx. It lives in *social* groups and its *habitat* is the plains. The cheetah has a *life span* of eight years. Its common *calls* are the hiss and purr, and its maximum *weight is* 125 pounds.

The **bobcat's** *life span* in its jungle *habitat* is six years. Its *social behavior* is defined by its solitary nature. The bobcat belongs to the *genus* Lynx, its primary *calls* are the hiss and purr, and its maximum *weight* is 30 pounds.

Fig. 2 Simple signaled wildcat text

An *extracted text* display handles the excess information problem found in the signaled text by physically extracting the signaled information from the larger text (Fig. 3). Our extracted text contained only the signaled information; the remaining text information was deleted. The extracted text was physically patterned after the signaled text such that the extracted segments appeared in the same locations as in the signaled text. Although this made an unusual and unlikely display, it did eliminate the excess information found in standard and signaled text, and it permitted us to investigate extraction independent of text signals and text reorganization. Still, text extraction alone does little to address the intervening-information problem. Maintaining the information's physical location means that potentially related information is still separated in space.

Displays intended to increase a reader's ability to recognize existing relationships by altering information's physical location have been developed. The present study examines two such displays—outlines and matrices—likely to help facilitate relationship identification.

An *outline* orders information in a hierarchical, list-like fashion (Kiewra et al. 1995; Robinson and Kiewra 1995). An outline representing the brief wildcat text appears in Fig. 4. The names of each wildcat are located corresponding to roman numerals I–IV; the categories (e.g., maximum weight), located directly below each cat, are marked by letters A–F; and the facts (e.g., lions weigh 400 pounds) are positioned directly beneath their corresponding category.

Unlike text, outlines extract important information. Unlike extracted text, outlines alter information's physical location and organize information hierarchically. Using the outline in Fig. 4, it is easy for the reader to see that the *lion belongs to the genus Panthera and has a maximum weight of 400 pounds,* and that *the cheetah belongs to the genus Acinonyx and*

Wildcats

tiger genus Panthera. call

roar. social behavior solitary habitat jungle.

life span 25 years weigh 450 pounds.

lion genus Panthera call

roar. plains habitat 25 year life span.

weighs 400 pounds. social group.

cheetah genus Acinonyx. social groups habitat

plains. life span eight years. calls hiss and

purr, weight 125 pounds.

bobcat's life span jungle habitat six years. social behavior

solitary genus Lynx, calls hiss

and purr, weight 30 pounds.

Fig. 3 Simple extracted wildcat text

has a maximum weight of 125 pounds. These two topical relationships are easily apparent. What is obscured, however, are the categorical relationships that exist within one category (e.g., *tigers and bobcats live in the jungle, whereas lions and cheetahs live on the plains*) or multiple categories such as the relationship between the wildcats' genus and their maximum weight—namely, that *cats from the genus Panthera weigh more than cats from other geniuses.* Although the outline's linear structure seemingly prevents readers from readily drawing relationships within categories, the matrix's structure seemingly overcomes this limitation.

The *matrix* is a two-dimensional cross-classification table that allows topics to be easily compared along one or more categories (e.g., Igo et al. 2008; Kiewra et al. 1999). For instance, the matrix presented in Table 1 allows the reader to read left-to-right along matrix rows, compare all four wildcats along the categories *habitat* and *social behavior,* and infer relationships such as: *solitary cats live in jungles, whereas plains cats live in groups.* That is why matrices work better than linear organizers for learning relationships (Kauffman and Kiewra 1998; Kiewra et al. 1999) and even solving real-world problems (Day 1988).

In summary, text, signaled text, extracted text, and outlines all appear to limit relationship learning relative to the matrix. The signaled text's organizational cues, the extracted text's extraction of important ideas, and the outline's extraction and linear organization of important information do not seem sufficient for helping students draw categorical relationships. In contrast, the matrix format seemingly allows readers to easily infer these relationships.

Wildcats

I. Tiger
   A. Genus
      1. Panthera
   B. Call
      1. Roar
   C. Weight
      1. 450 pounds
   D. Life span
      1. 25 years
   E. Habitat
      1. Jungle
   F. Social behavior
      1. Solitary

II. Lion
   A. Genus
      1. Panthera
   B. Call
      1. Roar
   C. Weight
      1. 400 pounds
   D. Life span
      1. 25 years
   E. Habitat
      1. Plains
   F. Social behavior
      1. Groups

III. Cheetah
   A. Genus
      1. Acinonyx
   B. Call
      1. Hiss and purr
   C. Weight
      1. 125 pounds
   D. Life span
      1. 8 years
   E. Habitat
      1. Plains
   F. Social behavior
      1. Groups

IV. Bobcat
   A. Genus
      1. Lynx
   B. Call
      1. Hiss and purr
   C. Weight
      1. 30 pounds
   D. Life span
      1. 6 years
   E. Habitat
      1. Jungle
   F. Social behavior
      1. Solitary

Fig. 4 Simple wildcat outline

Theoretical factors and related research

This study addresses the question, what types of displays best facilitate learning and why? A simple answer is that one display might present more information than another. If the displays are not "informationally equivalent" (Larkin and Simon 1987), then one display might have an advantage over the others. In the present study, the displays—text, signaled text, extracted text, outline, and matrix—were designed to maintain informational equivalence. Although they differed somewhat in their word counts, all reported the same topics, categories, and facts.

A second reason one display might be advantageous is that it might be more "computationally efficient" than another. Computational efficiency refers to how well a display allows a reader to locate important information and infer relationships. According to Larkin and Simon (1987), displays are computationally equivalent if they are first informationally equivalent and if any inference that can easily be drawn from one display can equally and easily be drawn from the other. A handful of researchers have theorized that a display's computational efficiency results from how well it signals, extracts, and localizes related information (e.g., Kiewra et al. 1999; Robinson and Kiewra 1995; Robinson and Skinner 1996). Each factor is described below.

Signaling

Signaling is a measure of how well a display cues information (Robinson and Skinner 1996; Titsworth and Kiewra 2004). Using the wildcat material as an example, topics such as tiger, categories such as weight, and facts such as 450 pounds can be signaled. The various displays seen earlier seem to differ in their signaling potential. For example, although the wildcat text is well organized, it does not signal information pertaining to *tiger, weight,* or the *tiger's weight.* The extracted text also lacks signals. In contrast, the signaled text, outline, and matrix all provide signals. All call attention to the topics, categories, and details. The signaled text uses bold-face type, italics, and underlining to signal information, whereas the outline and matrix spatially arrange the information so that topics, categories, and details are presented in clear and orderly locations.

To measure the independent effects of signaling, researchers must first control for extraction and localization effects. If displays differ in terms of signaling and extraction, for example, it is impossible to tell whether effects are due to signaling, extraction, or both. Studies that assess displays' signaling effects independent of extraction and localization are rare. One exception was a study by Robinson and Skinner (1996) that investigated how different displays—namely, text, outlines, and matrices—influenced how quickly and accurately students searched for information. Students read multiple-choice questions and then searched their assigned display for the correct response. Results indicated that students who searched a matrix or an outline located facts more quickly than those who searched a text, and that students who searched a matrix located relationships more quickly than those who searched an outline or a text. Robinson and Skinner concluded that the matrix's signaling potential improved students' ability to search because facts and relationships were more salient in the matrix as compared to the outline or text displays.

In the Robinson and Skinner (1996) study, the matrix contained signals whereas the text did not. Their study was confounded, however, because the matrix also extracted and localized important information whereas the text did not. This made it impossible to determine how the matrix's signaling potential impacted learning independent of extraction and localization. Additionally, students studying the outline and matrix should perform equivalently if signaling alone is at work because both displays contain comparable signals. However, because the matrix facilitated faster and more accurate responses as compared to the outline, observed differences must be due to other factors such as localization or extraction. Had displays been equivalent in terms of localization and extraction, then results could be attributed to signaling alone.

One means of signaling text is by inserting typographical cues, such as boldface type and underlining that signal the text's structure and content (see Mayer 2002 for a description of typographical cues). Lorch (1989) reported that reading text with typographical cues increased readers' memory for signaled content beyond reading text without typographical cues. Kiewra et al. (1999), in contrast, found that students who studied a signaled text performed no better on relationship tests than students who studied the text alone. Because the text and signaled text were equivalent in terms of extraction and localization, it appears that text signals have little impact on students' relationship learning.

Extraction

The second characteristic of computationally efficient displays is the degree to which they extract information. Extraction is the process of physically removing important content from intervening information. Extraction is useful because it allows readers to focus attention on key information and reduce the amount of information they must process in working memory. The displays investigated in the present study offer different degrees of extraction.

Neither the text nor signaled text extracts information for the reader. By their very nature, these displays embed facts and relationships within blocks of text, forcing the reader to sift through less important information in search of pertinent facts and relationships. In contrast to the text and signaled text, the extracted text, outline, and matrix displays extract information. They remove the most important text information from the less important information.

Measuring the independent effects of extraction depends on first controlling for signaling and localization. The I–IV-experiment study by Kiewra et al. (1995) accomplished this by investigating the benefits of supplementing text with outlines or matrices. These experiments controlled for signaling by comparing the achievement of students who reviewed (a) a research article without signaling, (b) the same article with important information signaled using underlining, (c) the article plus matrices, or (d) the article plus outlines. The matrix, outline, and underlined article signaled identical information. Following a 45-min study period, students were asked to recall facts and relationships. Results indicated that those who studied matrices or outlines recalled more information than those who studied the article with information underlined, who, in turn, recalled more (although not statistically) than those who studied the article alone. Because students who studied the matrices or outlines outperformed students who studied the signaled article, this study rejects the process of signaling alone and supports the process of extracting information from printed material.

The extraction research presented here (Kiewra et al. 1995) does an adequate job of controlling for possible signaling effects by including an underlined text that highlights important information. Because each display localized information differently, however, the study failed to separate extraction effects from localization effects. To assess how extraction impacts learning independent of signaling and localization, it is important that researchers devise a method for controlling localization and signaling while manipulating extraction. We do so in this study by introducing an extracted text.

Localization

The third characteristic of computationally efficient displays is localization, which refers to how close together similar information is placed on the printed page (Larkin and Simon 1987). We believe that two distinct types of localization exist: topical and categorical. First, displays that present all information about one topic followed by information about the next topic possess topical localization. For example, the wildcat outline in Fig. 4 presents all the information about tigers, followed by all the information about lions, cheetahs, and bobcats, respectively. Localizing information about a single topic might help the reader learn facts or relationships pertaining to that topic, but limit the learning of relationships across topics.

Categorical localization refers to how close together information spanning the same category (such as the wildcats' habitats) is placed on the printed page. Consider how the wildcat matrix in Table 1 localizes each wildcat habitat along the same row, whereas the outline (Fig. 4) separates this information in four distinct locations over two columns. Categorical localization allows the reader to compare topics easily across a single category such as call, and across multiple categories such as call and weight. For readers to see the interconnections among many ideas, or develop the “big picture,” both topical and categorical localization appear necessary.

In terms of localization, the text, signaled text, extracted text, and outline possess topical localization only. Information about each wildcat is located directly beneath its corresponding topic. Unfortunately, displays of this type separate similar categorical information (Kiewra et al. 1999). For example, if asked which wildcat has the shortest life span, a student studying the wildcat outline would have to locate the facts pertaining to each wildcat’s life span from the four distinct sections of the outline, hold each fact in working memory, then compare life spans to devise a response (Robinson and Skinner 1996). Doing so is clearly a time-consuming, effortful, and mistake-prone task. In contrast, the matrix localizes topic and category information using its two-dimensional structure (Kiewra et al. 1999). This structure allows the reader to easily locate a discrete fact or find any number of relationships within a topic or a category. For example, because each wildcat life span is located on the same row of the matrix, it is easy for the student to look across topics and determine that the bobcat has the shortest life span. Little effort is needed to accomplish this task.

To measure the independent effects of localization, the effects of signaling and extraction must be controlled. Studies that compare outlines and matrices do this because both displays extract information from the text and use equivalent signals. Several studies have compared outlines and matrices and found matrices superior to outlines particularly for relationship learning (Kauffman 2004; Kiewra et al. 1988; Kiewra, et al. 1991; Robinson and Kiewra 1995; Robinson and Schraw 1994).

Research comparing the outline and matrix generally supports the matrix as a more effective display. What research commonly fails to do, however, is identify what make the matrix so effective. Some might argue that the matrix’s advantage over outlines lies solely in its ability to present all the information in a smaller, more compact space (Kauffman et al. 2004). Although the matrix does hold this spatial advantage, that alone cannot explain its superiority. A study by Kiewra et al. (1999) suggests that the matrix’s advantage is not due solely to its spatial organization, but *how* the information is localized in the matrix. The ordering of matrix topics and categories affects relationship learning.

Summary

In summary, we have theorized that computational efficiency—the ability to locate information and infer relationships (Larkin and Simon 1987)—is the sum of a display’s ability to signal, extract, and localize related information. If so, the computational efficiency of the five displays investigated here increases linearly from text, signaled text, extracted text, outline, and matrix as shown in Table 2. Computational efficiency increases in this manner because the text and extracted text fail to signal information whereas the

Table 2 Levels of signaling, extraction, localization, and overall computational efficiency of displays

	Indexing	Extraction	Localization	Computational efficiency
Standard text	No	No	Topical	Very low
Signaled text	Yes	No	Topical	Low
Extracted text	No	Yes	Topical	Low
Outline	Yes	Yes	Topical	Moderate
Matrix	Yes	Yes	Topical and categorical	High

signaled text, outline, and matrix do signal information; the text and signaled text do not extract information but the extracted text, outline, and matrix do; finally, whereas text, signaled text, extracted text, and outlines provide topical localization at best, the matrix's two-dimensional structure affords topical and categorical localization. Research reviewed in this section suggests that various displays differ with regard to signaling, extraction, and localization and that those factors might explain the relative advantages some displays hold over other displays. Unfortunately, most of the research has not examined these theoretical factors independently. The present study attempts to do just that. Next, we recount the study's purpose and offer predictions.

Study purpose and predictions

The purpose of this study was to test the relative influences of signaling, extraction, and localization on the computational efficiency of displays. Although previous studies have interpreted research findings in terms of signaling, extraction, and localization, to date, researchers have not directly tested each construct's influence independently. In the present study, we accomplished this by developing unique displays such as an extracted text as well as traditional outline and matrix displays containing typographical signals. These displays allowed us to test each theoretical construct's relative contribution to students' learning.

To understand the unique contributions of signaling, extraction, and localization more fully, we bucked or followed certain conventions introduced in other display research. First, we used a lengthy and factual text. With a few exceptions (Igo and Kiewra 2007; Kiewra et al. 1999; Kiewra et al. 1995; Kiewra et al. 1994; Robinson et al. 2006; Robinson and Kiewra 1995), many of the previous display studies used relatively short text often containing fictitious information. The present study used a 2,000-word text containing factual information about six species of wildcats. Having students study a longer, fact-based text more closely resembles the length and quality of materials students study in school.

Second, tests measured facts and relationships. Because various displays might particularly facilitate fact or relationship learning, we followed the precedent set by some previous adjunct display studies and measured both learning outcomes (Jairam and Kiewra in press; Kauffman 2004).

Third, testing occurred immediately after a 15-min study period, following a 1-week delay, and after students had an opportunity to "relearn" the material with additional study. Past display research oftentimes ignored the delayed effects of studying displays (Kiewra et al. 1995) or limited the delay to only a few days (Robinson and Kiewra 1995). Only one other study (Kiewra et al. 1999) examined how well information is relearned from a display. Certain text aids might have their most profound effects following a delay or when used as tools for relearning.

Fourth, to ensure that student learning was influenced by displays and not other factors, we eliminated student note taking, a practice common in previous adjunct display research (e.g., Kiewra et al. 1999; Robinson and Kiewra 1995).

Last, we joined other investigators who examined students' attitudes toward the displays (e.g., Igo et al. 2009; Igo and Kiewra 2007; Kauffman 2004). Students liking and valuing a display seems important in motivating them to use displays on their own.

Our predictions pertain to the independent and relative value of signaling, extraction, and localization. First, we predicted that signaling and extraction alone would have mild effects on fact learning and no substantial effect on relationship learning. Both signaling and extraction highlight key information either by emphasizing it (signaling) or isolating it from less important information (extraction). As such, displays that signal or extract important information should boost attention to designated ideas and increase fact learning but should not prompt the integration of ideas nor increase relationship learning. Outlines and matrices should facilitate relationship learning because both promote localization of ideas. Because matrices promote both topical and categorical localization, whereas outlines promote topical localization only, matrices should produce the best relationship learning.

Experiment 1

Experiment 1 examined the independent and relative effects of signaling, extraction, and localization on learning from displays. Experiment 2 was an extension of Experiment 1 and looked solely at varying localization within matrix displays.

Method

Participants and design

Seventy two undergraduate students (20 males and 52 females) from a large Midwestern university were assigned randomly to one cell of a 4×2 design. The first factor was displays. Students studied a passage, an extracted passage, an outline, or a matrix. The second factor was the presence or absence of signaling cues. Groups did not differ with respect to gender, age ($M = 22.74$), or class standing (most were juniors).

Materials

Materials included a pre-experimental survey, the four types of displays (with or without signals), an audio-taped lecture, three tests designed to measure recognition of facts and relationships, and a post-experimental questionnaire.

The pre-experimental survey obtained demographic information from participants such as gender, age, and class standing. It also ascertained participants' self reported levels of prior knowledge regarding biology in general and wildcats in particular. General biological knowledge was determined by having students circle a number corresponding to the number of biology courses previously taken. Also, students indicated how much they knew about biological classification by responding on a four-point Likert scale. To find out what students knew about wildcats in particular, students answered the question, "How much do you know about wildcats?" using the same Likert scale.

The displays were similar in content and structure to the wildcat displays found in the introduction but were far more comprehensive. The standard text contained 1998 words and was typed single space in two columns across five pages. The text was adapted from a number of wildcat resources by the researchers. It described six different wildcats in this order: tiger, lion, jaguar, leopard, cheetah, and bobcat. Each wildcat was described in the same order along 13 categories that fit into three major classifications: physical features (genus, call, weight, coat, and distinctive characteristic), life style (habitat, range, social behavior, and life span), and hunting behavior (what, when, method, and frequency). The text contained 78 important facts pertaining to the wildcats and their categories. An example fact is, “the lion’s habitat is the plains.” The standard text was designed to read as clearly as possible. For example, by organizing the content relative to topic (e.g., tiger), major classification (e.g., lifestyle), and category (e.g., range), it provided the reader information about the text’s hierarchical macrostructure, a key feature of a well-designed text (Van Djik and Kintch 1983). Although we certainly could have created an alternate text organized by category instead of topic or one that overtly expressed relationships among wildcats, ours simulated the majority of texts that adopt a topic-by-topic rather than comparative structure (Jonassen et al. 1993).

In the extracted text, the topics, categories, and facts were physically removed from the standard text and were presented on the printed page in their original locations. In essence, all non-topic, non-category, and non-fact information was deleted from the standard text leaving all 78 facts presented in a total of 351 words. Although the extracted text is not a traditional or recommended display, it was created to assess how extraction independently influences fact and relationship learning.

The outline and matrix displays contained the same information as the extracted text. The outline display organized information in two columns across three pages. The columns on a page were not aligned. For example, information about the tiger and lion’s habitats were on different horizontal lines making their comparison across categories less than ideal. The wildcats’ names appeared as roman numerals I–VI. Beneath each roman numeral, the three major classifications appeared as capital letters A–C. The 13 categories appeared as numbers 1–5, and important facts were assigned lower case letters and subsumed beneath their respective categories. A total of 367 words appeared in the outline.

The matrix display was a two-dimensional classification table appearing on a single page. It listed the wildcats’ topics along the top row, major classifications and categories down the left-most column, and facts intersecting topics and categories within their matrix cells. It contained 244 words.

A second set of displays was constructed. They were identical to the first (standard, extracted, outline, and matrix) except for the addition of signals. Specifically, topic names (e.g., tiger) and major classification names (e.g., life style) were typed in bold face, category names (e.g., social behavior) were italicized, and facts (e.g., solitary) were underlined. The signaled displays were constructed to highlight the 78 wildcat facts and their organization and to assess the independent effects of signaling. By comparing a standard display with its signaled version, we were able to control for extraction and localization and thus assess signaling independently.

A 13-min audio-taped lecture identical to the text was developed. The lecture was presented at approximately 170 words per minute. The purpose of the lecture was to expose all participants uniformly to the complete wildcat material before they studied their group-specific displays.

Three multiple-choice tests were developed. The *fact* test contained 20 items that measured students’ ability to recognize discrete facts about wildcats. An example fact test question is, “What is the maximum weight of the tiger?” The *local relationship* test (Winn 1991) was a 20-item multiple-choice test that assessed participants’ ability to recognize wildcat relationships along a single category. An example local relationship test question is, “Which two cats hiss and purr?” The *global relationship* test (Winn 1991) contained 10 multiple-choice items that assessed how well students could recognize wildcat relationships along two or more categories. An example global relationship test question is, “What is the relationship between call and weight?”

A post-experimental questionnaire assessed students’ attitudes about their study materials. This questionnaire asked students to rate the (a) ease of reviewing their displays, (b) length of time they needed to study, (c) how prepared they felt for the tests, and (d) their interest in using similar displays in the future. Students responded to each question using a four-point Likert scale, with responses 1–4 corresponding to “strongly agree,” “agree,” “disagree,” and “strongly disagree,” respectively.

Procedure

Experiment 1 occurred in seven phases: survey, acquisition, immediate testing, delayed testing, relearning, retesting, and questionnaire. During phase one, participants completed the demographic survey. Phase two, the acquisition phase, began with students listening to the audio taped lecture while reading along with the full non-signaled text. This was done to ensure that all students were exposed equally, both aurally and visually, to the complete wildcat materials before studying their group-specific displays. Research confirms that simultaneous presentation of information through multiple channels is best (Moreno and Mayer 2002). Following the lecture, students were required to place the full text in a folder and to remove their study materials. Participants reviewed their study materials (standard text, extracted text, outline, or matrix, with or without signals) for 15 min, without the benefit of note taking. Students were told their studying was in preparation for three tests that assessed how well they learned facts and relationships. Examples of each test type were provided.

Phase three, immediate testing, followed the acquisition phase. Participants completed the global relationship, local relationship, and fact tests in that order at their own pace. Tests were administered in that order to limit the likelihood that a test would influence responses to subsequent tests. Upon completion, testing materials were collected and students were dismissed.

One week later, during the fourth phase, delayed testing, participants were tested without warning using the identical tests and same testing procedure from a week ago. Immediately following, during phase five, participants were given the opportunity to “relearn” the content using their same displays for 10 min, “in preparation for the same tests.” The same three tests were re-administered in the same way in phase six to assess the savings potential of the study materials. During phase seven, students completed the post-experimental questionnaire. Following debriefing, materials were collected and students were excused.

Results

Results pertain to students’ prior knowledge, test performance, and attitudes about their displays.

Prior knowledge

Each item from the prior knowledge questionnaire was analyzed separately using an ANOVA procedure to determine whether display groups differed in terms of (a) biological coursework, (b) knowledge of biological classification, or (c) knowledge of wildcats. No significant differences occurred among groups on these three measures, all F’s (3, 68), $p < 1.05$. Participants generally had little prior knowledge regarding the topic. On average, students had taken one biology course, knew “next to nothing” about biological classification, and knew “a little” about wildcats.

Test performance

Tables 3, 4, and 5 present means and standard deviations for display groups with and without signals and in total on each administration of the global relationship, local relationship, and fact tests, respectively. Separate 4 (displays) × 2 (signals) × 3 (test administration) ANOVA procedures were conducted for the global relationship, local relationship, and fact tests (with test administration as the within-subjects factor). Significant main effects ($p < 0.05$) were followed up with Tukey post hoc analyses to determine significant differences among groups. Results pertaining to signals revealed no main effects or interactions (all $F < 1.2$). Consequently, the following sections report results pertaining to the four display groups without regard for signals.

Global relationship test

The ANOVA revealed two main effects and an interaction. First, a main effect for test administration was revealed, F (2, 142) = 12.10; $p < 0.01$, MSe = 9.75. As seen in the bottom row of Table 3, students performed better on the immediate test and the savings test than on the delayed test.

Second, a main effect for displays was observed, F (3, 64) = 21.29, $p < 0.01$, MSe = 71.52. As seen in the right-most column of Table 3, students who studied the outline outperformed those who studied the text and those who studied the extracted text. Additionally, students who studied the matrix outperformed those who studied the text, extracted text, and outline.

Table 3 Means and standard deviations for display groups on each administration of the global relationship test in Experiment 1

Displays	Immediate		Delayed		Savings		Total	
	M	SD	*M*	SD	*M*	SD	*M*	SD
Text								
Signaled ($n = 9$)	7.00	1.58	5.78	1.48	6.88	1.76	6.72	1.16
Not signaled ($n = 9$)	7.11	.93	6.33	1.32	7.22	1.39		
Total ($n = 18$)	7.06	1.26	6.06	1.13	7.06	1.55		
Extracted text								
Signaled ($n = 9$)	5.89	1.17	6.22	1.30	6.44	2.19	6.37	1.14
Not signaled ($n = 9$)	6.78	1.56	6.33	1.00	6.56	1.01		
Total ($n = 18$)	6.33	1.44	6.28	1.13	6.50	1.65		
Outline								
Signaled ($n = 9$)	7.67	1.41	7.33	1.12	7.44	1.01	7.67	1.17
Not signaled ($n = 9$)	7.67	1.41	7.67	1.73	8.22	1.20		
Total ($n = 18$)	7.67	1.37	7.67	1.42	7.83	1.15		
Matrix								
Signaled ($n = 9$)	9.11	.60	8.44	1.13	9.33	.70	8.94	0.55
Not signaled ($n = 9$)	9.11	.78	8.00	1.00	9.67	.50		
Total ($n = 18$)	9.11	.68	8.22	1.06	9.50	.62		
Total ($n = 72$)	7.54	1.57	7.01	1.52	7.72	1.71	7.43	1.42

Table 4 Means and standard deviations for display groups on each administration of the local relationship test in Experiment 1

Displays	Immediate		Delayed		Savings		Total	
	M	SD	*M*	SD	*M*	SD	*M*	SD
Text								
Signaled ($n = 9$)	14.33	1.66	11.89	1.97	14.89	2.32	13.94	1.93
Not signaled ($n = 9$)	13.56	2.00	13.11	2.62	15.89	2.85		
Total ($n = 18$)	13.94	1.87	12.50	2.33	15.39	2.57		
Extracted text								
Signaled ($n = 9$)	13.78	2.69	12.89	2.47	14.78	3.49	13.35	2.24
Not signaled ($n = 9$)	12.78	1.98	12.33	1.58	13.56	3.50		
Total ($n = 18$)	13.28	2.35	12.61	2.03	14.17	3.45		
Outline								
Signaled ($n = 9$)	15.78	1.72	13.89	3.18	15.78	4.12	15.11	3.10
Not signaled ($n = 9$)	15.22	3.23	14.11	4.01	15.89	4.51		
Total ($n = 18$)	15.50	2.53	14.00	3.51	15.83	4.19		
Matrix								
Signaled ($n = 9$)	15.50	.78	15.56	1.24	17.44	1.13	17.52	1.18
Not signaled ($n = 9$)	18.89	1.05	16.22	2.54	19.11	.93		
Total ($n = 18$)	18.39	1.04	15.89	1.97	18.28	1.32		
Total ($n = 72$)	15.28	2.80	13.75	2.84	14.67	3.44	14.98	2.71

Table 5 Means and standard deviations for display groups on each administration of the fact test in Experiment 1

Displays	Immediate		Delayed		Savings		Total	
	M	SD	*M*	SD	*M*	SD	*M*	SD
Text								
Signaled ($n = 9$)	11.33	2.74	10.22	2.43	13.55	2.40	12.17	2.50
Not signaled ($n = 9$)	13.22	2.99	9.67	3.81	15.00	2.65		
Total ($n = 18$)	12.28	2.97	9.94	3.11	14.28	2.56		
Extracted text								
Signaled ($n = 9$)	13.89	2.42	10.33	3.78	12.56	3.05	12.00	2.65
Not signaled ($n = 9$)	12.33	4.18	10.00	2.50	12.89	2.37		
Total ($n = 18$)	13.11	3.41	10.17	3.11	12.72	2.65		
Outline								
Signaled ($n = 9$)	15.33	2.06	11.78	4.18	14.56	4.42	13.94	3.55
Not signaled ($n = 9$)	15.67	2.65	11.44	4.56	14.89	5.49		
Total ($n = 18$)	15.50	2.31	11.61	4.24	14.72	4.84		
Matrix								
Signaled ($n = 9$)	16.33	1.32	13.44	1.94	16.33	1.66	15.72	1.40
Not signaled ($n = 9$)	17.00	1.66	13.67	2.00	17.56	1.51		
Total ($n = 18$)	16.67	1.50	13.56	1.92	16.94	1.66		
Total ($n = 72$)	14.39	3.14	11.32	3.46	14.67	3.44	13.46	3.00

Finally, a modest test administration X displays interaction was observed, F (6, 128) = 2.01, $p < 0.05$, MSe = 1.57. As seen in the internal cells of Table 3, test scores dropped more precipitously from immediate to delayed testing for the text group than for the other display groups. In fact, the text group had the third-best score on the immediate and savings tests but the worst delayed score.

Local relationship test

The ANOVA revealed two main effects. First, a main effect for test administration was observed F (2, 142) = 33.87, $p < 0.01$, MSe = 89.24. As seen in the bottom row of Table 4, participants performed better on the immediate test and savings test than on the delayed test.

A main effect was also revealed for displays, F (3, 64) = 12.00, $p < 0.01$, MSe = 183.32. As seen in the right-most column of Table 4, students who studied the matrix outperformed those who studied the text, extracted text, and outline. Additionally, students who studied the outline outperformed those who studied the extracted text.

Fact test

Two main effects and an interaction were observed. First, a main effect for test administration was observed, F (2, 128) = 78.67, $p < 0.01$, MSe = 248.43. As seen in the bottom row of Table 5, students performed better on the immediate and savings tests than the delayed test.

A main effect was also observed for displays, F (3, 64) = 7.55, $p < 0.01$, MSe = 7.55. As seen in Table 5's right-most column, students who studied the outline outperformed those who studied the text and those who studied the extracted text. Additionally, students who studied the matrix outperformed those who studied all other displays.

A displays X test administration interaction was also observed, F (6, 128) = 2.35, $p < 0.05$, MSe = 7.41. As seen in the internal cells of Table 5, students who studied the extracted text outperformed those who studied the text on the immediate and delayed tests, but not on the savings test where the text group's score rose markedly, although it still fell behind the savings scores of the outline and matrix groups.

Attitudes about study materials

Table 6 presents means and standard deviations for each display group on four items from the post-experiment survey. Separate 4 (displays) × 2 (signals) ANOVAs were conducted on the mean responses to each question. Significant main effects ($p < 0.05$) were followed up with Tukey post hoc analyses to determine significant differences among groups.

Table 6 Means and standard deviations for each post-experiment survey question in Experiment 1

Displays	Ease		Time		Prepare		Reuse	
	M	SD	*M*	SD	*M*	SD	*M*	SD
Text (n = 18)	2.56	0.98	2.28	0.67	2.50	0.79	2.78	0.94
Extracted text (n = 18)	2.89	0.90	2.39	0.61	2.44	0.62	3.28	0.67
Outline (n = 18)	2.00	0.51	2.11	0.58	1.67	0.59	1.94	1.00
Matrix (n = 18)	1.56	0.51	1.78	0.73	1.50	0.51	1.56	0.62

Survey Item 1 revealed a main effect for displays, F (3, 64) = 8.22, $p < 0.01$, MSe = 0.76. As seen in the "Ease" column of Table 6, those who studied the matrix reported it was easier to review their materials than those who studied either the text or the extracted text. Similarly, students who studied the outline reported it was easier to review their materials than those who studied either the text or the extracted text.

Survey Item 2 revealed a main effect for displays, F (3, 64) = 3.15, $p < 0.05$, MSe = 1.28. As seen in the "Time" column in Table 6, those who studied the matrix were more likely to report they had enough time to study than those who studied the extracted text or the text. No difference existed between those who studied the outline and those who studied other displays.

Survey Item 3 assessed participants' belief about how well the materials helped them prepare for the tests. A main effect for displays was again revealed, F (3, 64) = 12.21, $p < 0.01$, MSe = 4.83. As seen in the "Prepare" column in Table 6, students who studied the matrix reported feeling more prepared than those who studied the extracted text and those who studied the text. Similarly, students who studied the outline reported feeling more prepared that those who studied the text and those who studied the extracted text.

Survey Item 4 assessed participants' satisfaction with their displays as judged by participants' willingness to reuse displays like this in the future. Once again, a main effect for displays was observed, F (3, 64) = 16.85, $p < 0.01$, MSe = 11.00. As seen in the "Reuse" column in Table 6, the matrix group reported greater satisfaction than the text group and extracted text group. Participants who studied the outline reported greater satisfaction than those who studied the text and those who studied the extracted text.

Discussion

Experiment 1 examined the independent and relative effects of signaling, extraction, and localization. It was thought that displays that signal, extract, or localize information might facilitate learning best. This discussion centers on those three factors.

Highlighting important topics, categories, and facts alone appears insufficient for learning

Contrary to our prediction, typographical signals in the form of bold-face type, italics, and underlining did not facilitate test performance. We had predicted that signaling might raise fact test performance because the signals highlight important topics, categories, and the facts pertaining to the intersection of those topics and categories. Results from fact and relationship tests showed no main effect favoring displays with signals over displays without signals. In no case, over three administrations of three different tests, did signals help despite the fact that sample size (n = 36 per group) was adequate to detect even small differences. On the contrary, students who studied the signaled materials actually scored somewhat lower than those who studied the non-signaled materials on the majority of tests as shown in the internal cells in Tables 3, 4 and 5.

It might also be argued that outlines and matrices inherently contain signals by virtue of their structure, and that adding typographical cues like italics and boldface type is redundant. Although the redundancy principle might explain why signaling did not improve outlines and matrices, it does not explain why signaling failed to improve text learning. Oddly, in seven of nine cases, the non-signaled text led to somewhat higher test performance than signaled text as shown in the top-most rows in Tables 3, 4 and 5.

Our finding that signals are not helpful is consistent with that of Kiewra et al. (1999) but largely at odds with studies reviewed by Lorch (1989). Perhaps the discrepant findings result from differences in the cues' functions. In most studies, typographical signals, like those used here, signal important information for selection. In this study (and in Kiewra et al.) the signals cued the text's organization by highlighting and distinguishing topics, major classifications, categories, and facts.

Differences in signal function perhaps created another difference between the present study and those reviewed by Lorch (1989). Because signals in our study served an organizational function more signals were used than if signals served a selection function. In fact, nearly 20% of the text contained signals. Lorch (1989) found that signaling was generally ineffective when more than 17% of material was signaled. When signaling exceeds that amount, signals lose their distinctiveness and ability to distinguish key ideas (Schmidt 1991).

Because the signaled displays contained a large percentage of signals, it is understandable that signals did not serve a selection function and boost fact learning as might ordinarily be the case. Why signals that are organizational in nature did not serve an organization function and boost relationship learning is also understandable. Computational efficiency (Larkin and Simon 1987) cannot be achieved by nominally classifying information as a topic or category. Instead, information must be physically organized or "localized" for relationship learning to occur. That signals did not boost relationship learning is an important addition to the signaling literature. Past studies usually examine signal effectiveness relative to fact learning only (Lorch 1989).

Finally, it is possible that no signaling effect was found because text signals tend to aid low achieving students more than high achieving students (Meyer et al. 1980). General achievement was not measured in the present study and students were generally high achievers—juniors and seniors enrolled at a selective university.

Extraction alone is insufficient for learning

Contrary to our prediction, extracting important ideas from text did not facilitate test performance. We had predicted that extracting key information about topics, categories, and the facts pertaining to the intersection of those topics and categories might raise fact performance because students' attention is focused on this isolated information rather than divided between this important information and the less important information found in the complete text.

In no case, on neither fact nor relationship tests, did students studying the extracted text outperform students studying the complete text. In fact, a casual inspection of the rightmost columns of Tables 3, 4 and 5 shows a modest advantage for studying standard text over extracted text. The test occasion by displays interaction results also suggest that the complete text is especially more effective than the extracted text upon immediate testing and following an opportunity to relearn.

In conclusion, simply extracting key information from text appears to be insufficient for boosting fact or relationship learning. As is next shown, key information must not only be extracted but also localized in order to boost learning.

Localization bolsters fact and relationship learning

In support of our prediction, localization was an important factor in determining a display's benefit. The most direct test of localization came from comparing the performance of outline and matrix studiers versus the extracted text studiers. All these displays extracted important ideas, but only the outline and matrix displays localized—or organized—those ideas.

Students studying the outline or matrix outperformed those studying the extracted text on the global relationship test, local relationship test, and fact test. This is convincing evidence that it is not enough to extract or select-out key information from text. The extracted information must be meaningfully organized.

Findings also confirmed the prediction that the matrix localizes information better than the outline. On all three tests—global relationship, local relationship, and fact—the matrix group outperformed the outline group. Our explanation is that outlines provide within-topic, or topical, localization. They organize information by topics such as tiger or lion. Matrices, on the other hand, provide both within-topic and within-category localization. The wildcat matrix organized information topically by wildcat (e.g., tiger and lion), but also categorically (e.g., by call and weight). To study all the wildcat calls, the reader simply looks across the single matrix row designated "call." All the "call" information appears in close proximity in the matrix, whereas that same information appears in six different and separated locations in the outline. The close proximity of within-category information helps the matrix studier learn local relationships within one category and global relationships across two or more categories.

Display benefits hold across testing occasions

Generally speaking, test results showed that studying a matrix was superior to studying an outline and both were superior to studying text or extracted text. Other than a couple mild interactions, these findings were consistent across immediate, delayed, and savings tests. This is important because it shows the superiority of matrices, with their topical and categorical localization, across diverse testing conditions and shows that even under the best of testing conditions (after relearning) studying standard text or even extracted text is relatively ineffective. Consider that when tested at "relearning," students had already studied, been tested twice previously, and had just studied a second time now knowing the very questions that would appear during the third testing session. Despite all this, students studying text or extracted text still performed relatively poorly compared to outline and matrix studiers and performed poorly in an absolute sense as well. On average, the two text groups scored about 68% on the global relationship savings test, and 67% on the fact savings test. Despite two studying opportunities and two test opportunities, studying text just was not effective even under ideal learning and testing conditions. Again, the consistency of findings across testing conditions underscores the benefits of matrices and the limitations of text.

Student satisfaction mirrors display effectiveness

Test performance results confirmed that matrices are better study devices than outlines which, in turn, are better than extracted text or standard text. Students' perceived satisfaction with displays mirrored test results.

Students studying the matrix consistently preferred this display to those who studied text or extracted text. The matrix studiers rated their materials easier to use, more efficient, more effective, and more desirable for future use. Outline studiers also rated their displays easier to use, more efficient, and more desirable than did text or extracted text studiers.

Although perceptions between matrix and outline users were not reliably different, inspection of Table 6 shows that the matrix users consistently rated their displays better than did outline users. These results are in line with previous research by Kauffman (1998), Kauffman and Kiewra (1998), and Robinson and Kiewra (1995) showing that students who study a matrix are more satisfied than students who study outlines or text. Overall, these results show that students are sensitive to the relative benefits of displays used for studying and prefer and value those that do the most for them.

Experiment 2

Experiment 1 showed the superiority of the matrix display over other displays and confirmed that its superiority is due to the presence of both topical (within-topic) and categorical (within-category) localization. Experiment 2 was conducted to see if the story ends there. On one hand, a matrix, any matrix, might be considered effective because it imposes topical and categorical localization. On the other hand, perhaps not all matrices are created equal. Perhaps altering a matrix's ordering of topics or categories enhances or diminishes localization. Experiment 2 examined the performance effects of studying matrices considered high in topical localization, high in categorical localization, high in both, or weak in both. The results should clarify the relative contribution of topical and categorical localization and provide practical implications for matrix construction.

Method

Participants and design

Fifty four additional undergraduate students from the same subject-pool as in Experiment 1 were assigned randomly to one cell of a 2 × 2 design. (One student's data were omitted, however, because of that student's noncompliance.) The first factor was topical organization. Students studied a matrix with either logically or randomly organized topics. The second factor was categorical organization. Students studied a matrix with either logically or randomly organized categories. Results from a pre-experimental survey (identical to the one used in Experiment 1) showed that, on average, participants were 21 years old, of primarily junior standing, had minimal coursework in biology, and had little knowledge of biological classification or wildcats.

Materials

Four matrix displays adapted from the wildcat material used in Experiment 1 were constructed. All contained the same six topics (wildcat names) across the top, six categories (call, weight, life span, habitat, social behavior, and range) down the left side, and 36 facts, each within a cell corresponding to its topic and category. Thus, the four matrix displays were informationally equivalent. The four matrices differed in that topics and categories were ordered either logically or randomly. The logical ordering of topics (as seen in Matrix 1 and 2 in the Appendix) ordered the wildcats from biggest to smallest when reading from left to right. Notice in Matrix 1 and 2 that the first two cats (tiger and lion) are heavy, the next two cats (jaguar and leopard) are moderate in weight, and the last two cats (cheetah and bobcat) are light in weight. This logical ordering of topics also shows a natural decline in life span and in the voracity of the cats' calls as the matrices are read from left to right. In Matrix 3 and 4, as shown in the Appendix, the ordering of topics (wildcat names) was random making it difficult to notice the progressive decline in weight, life span, and call from the largest cats (tiger and lion), to the medium sized cats (jaguar and leopard), to the smallest cats (cheetah and bobcat).

Categories were also ordered logically (in Matrix 1 and 3) or randomly (in Matrix 2 and 4). Matrices with logically ordered categories placed the categories call, weight, and life span in succession, and the categories habitat, social behavior, and range in succession. The categories call, weight, and life span should be grouped because there is a logical relationship among these variables. Cats that roar are heavy and live long lives. Cats that growl are moderate in weight and have medium life spans. Cats that purr are light in weight and have short life spans. Similarly, the categories habitat, social behavior, and range should be grouped because there is a logical relationship among these variables. Cats that live in the jungle are solitary and have confined ranges. Cats that live on the plains live in groups and have vast ranges.

In summary, the four matrices had either logically ordered topics (Matrix 1 and 2) or randomly ordered topics (Matrix 3 and 4) and either logically ordered categories (Matrix 1 and 3) or randomly ordered categories (Matrix 2 and 4). This means that Matrix 1 was the most logically ordered matrix overall and that Matrix 4 was ordered most randomly or illogically.

Three tests similar to those in Experiment 1 assessed participants' knowledge of wildcat facts and relationships. The *fact* test contained 18 fill-in-the-blank items assessing knowledge of explicitly presented facts. An example of a fact question is, "The jaguar's habitat is the____." The *local relationship* test contained 20 multiple-choice items assessing participants' knowledge of existing relationships among wildcats within a single category. An example of this type of question is, "Which two wildcats purr?" The *global relationship* test was a two-part test that measured participants' knowledge of relationships derived from multiple categories. Part One included an essay question asking students to describe any "overriding relationships" they saw in the materials. Part Two included six short-answer questions that assessed knowledge of specific overriding relationships. An example short-answer question is, "What is the relationship between wildcats' call and weight."

Procedure

Participants received folders containing their particular matrix display (high or low topical organization and high or low categorical organization) and test materials. Oral instructions were given informing all students that they would study a matrix display for six minutes and then take fact and relationship tests. Students reviewed their matrix display for six minutes without an opportunity to record additional notes. Students then turned in their study material and completed a five-minute distracter task to clear short-term memory and minimize rehearsal effects. Participants then completed the local relationship, global relationship, and fact tests in that order. They were given unlimited time but took approximately 7 min to complete each test. After completing the fact test, participants turned in their folders, were debriefed, and dismissed.

Results and discussion

Separate 2 × 2 ANOVAs were conducted for scores on the local relationship, global relationship, and fact tests. On the local relationship test, a main effect was observed for topical organization only, F (1, 50) = 4.44, $p < 0.05$, MSe = 8.44. Participants who studied a matrix with logically organized topics outperformed those who studied a matrix with randomly organized topics as seen in the bottom row of Table 7. That topical organization plays an important role in learning local relationships makes sense because placing similar topics close together decreases the amount of intervening information between related ideas within a matrix. Notice in the Appendix, for example, how easy it is to see "call" relationships across Row 1 in Matrix 1 with organized topics versus Row 1 of Matrix 3 with randomly organized topics.

Results from the global relationship test revealed a main effect only for topical organization as well, F (1, 50) = 6.79, $p < 0.05$, MSe = 127.29. Students who studied matrices with logically organized topics outperformed those who studied matrices with randomly organized topics as seen in the bottom row of Table 8. This finding confirms that topical organization is vital when learners must discern overarching relationships within multiple matrix rows. Notice in the Appendix, for example, how easy it is to discern that heavier wildcats have louder calls and live longer than lighter weight wildcats when examining the top three rows of Matrix 1 because the wildcats (the topics) are ordered from heaviest to lightest. In contrast, it is difficult to discern that same relationship in Matrix 3 because the topics (wildcats) are ordered randomly thereby separating the cats with similar weights, life spans, and calls.

Finally, there was only a main effect for topical organization on the fact test as well, F (1, 49) = 8.55, $p < 0.01$, MSe = 13.20. Once again, students who studied matrices with logically organized topics outperformed those who studied matrices with randomly

Table 7 Means (and standard deviations) for matrix groups on local relationship test in Experiment 2

Categories	Topics		
	Organized (n = 13)	Random (n = 13)	Total
Organized (n = 14)	19.00 (1.92)	16.92 (4.06)	18.04 (3.21)
Random (n = 13)	18.64 (2.10)	17.46 (3.36)	18.07 (2.79)
Total (n = 53)	18.82 (1.98)	17.15 (3.64)	18.06 (2.97)

Table 8 Means (and standard deviations) for matrix groups on the global relationship test in Experiment 2

Categories	Topics		
	Organized (n = 13)	Random (n = 13)	Total
Organized (n = 14)	34.50 (9.74)	21.92 (11.38)	28.69 (12.14)
Random (n = 13)	26.36 (12.41)	23.85 (11.44)	25.15 (11.80)
Total (n = 53)	30.43 (11.71)	22.92 (11.22)	26.89 (11.98)

Table 9 Means (and standard deviations) for matrix groups on the fact test in Experiment 2

Categories	Topics		
	Organized (n = 13)	Random (n = 13)	Total
Organized (n = 14)	14.93 (2.53)	10.25 (3.47)	12.77 (3.78)
Random (n = 13)	13.79 (4.15)	12.62 (4.15)	13.22 (4.12)
Total (n = 53)	14.36 (3.42)	11.48 (3.95)	13.00 (3.92)

organized topics as seen in the bottom row of Table 9. Because topical organization boosted relationship learning, it makes sense that it also boosted fact learning because knowledge about relationships also helps to remember facts. Learning the relationship that heavy cats roar helps in learning subordinate facts such as tigers are heavy and tigers roar.

Overall, the results of Experiment 2 confirm that not all matrices are created equal. It is most important to order the topics logically. When topics are logically organized, it is possible to examine information within a category (such as call) and learn local relationships such as "tigers and lions roar," or to examine information within multiple categories (such as call and weight) and learn global relationships such as "the heavier the cat, the louder its call." Fact learning is also boosted by creating logically organized topics. Our speculation is that relationship learning provides the basis for learning the subordinate facts inherent in relationships. When topics are not logically organized, as in Matrices 3 and 4 in the Appendix, then local and global relationships and their accompanying facts are more difficult to learn.

Although topical organization had a statistically significant effect on test performance, categorical organization did not. A closer inspection of the data, however, shows that categorical organization had a practically important effect. First, examine the means in the interior of Tables 7, 8 and 9. On all three tests, the highest performance was by the group studying the matrix containing both organized topics and organized categories. The combination of organized topics and categories seems to work best for learning facts and relationships. Second, inspect the right-most columns of Tables 7, 8 and 9 and note that category organization had its most pronounced effect on learning global relationships (Table 8). Cohen (1992) effect sizes confirm this pattern. Categorical organization effect sizes were 0.01 for local relationships, −0.11 for facts, but 0.30 for global relationships. Categorical organization's differential effects for the three performance tests make sense. Varying category organization should have minimal effects on learning isolated facts or relationships within a single category. Varying category organization, however, should affect learning global relationships because these are drawn across multiple categories. If related categories are separated by intervening categories, then global relationships should be more difficult to discern. For example, it seems easier to see the global relationship that heavier cats have longer life spans in Matrix 1 in the Appendix than in Matrix 2 because Matrix 1 localizes the categories of weight and life span. Those categories are adjacent in Matrix 1 but separated by three other categories in Matrix 2. Still, students might not ordinarily study by comparing matrix rows. A recent study (Jairam and Kiewra 2009) shows that highlighting related rows helps students attend to categorical organization and learn global relationships.

General discussion

Our purpose was to determine (a) what type of display works best for helping students learn facts and relationships and (b) why a display is effective. We found that a matrix display boosts fact and relationship learning more than a standard text, a signaled text, an extracted text, or an outline. The reason the matrix is superior is because it localizes related information better than other displays.

The matrix's two-dimensional structure allows studiers to look across a single matrix row (or category) and easily compare topics such as wildcats. For instance, all information about wildcats' call appear in the same matrix row in Table 1 making it easy to learn the local relationship that two cats (tiger and lion) roar and two cats (cheetah and bobcat) hiss and purr. This same information is separated in the other displays (Figs. 2, 3, 4). The matrix also permits studiers to look across multiple rows (or categories) and easily compare topics. For instance looking across Table 1's rows for call and weight, it is easy to learn the global relationship that heavier cats roar whereas lighter weight cats hiss and purr. The eight facts that comprise this global relationship are localized within adjacent matrix rows but are dispersed throughout the other displays.

Matrices work best because they provide better localization than other displays. However, not all matrices are created equal. The ordering of matrix topics and categories affects localization and, therefore, learning. Matrices work best when topics follow a natural order such as when wildcats are presented from heaviest to lightest or from longest to shortest life span. When the natural ordering of topics is varied, then students have difficulty learning local and global relationships and the facts that comprise those relationships. Varying the natural ordering of categories, however, has little effect on fact learning or local relationship learning but a modest effect on learning global relationships. Here is why. When topic order is varied the local relationships within each matrix row or category and the global relationships within multiple matrix rows or categories become obscured. Note the differences between Matrix 1 (with organized topics) and 3 (with random topics) in the Appendix. In Matrix 1, wildcats with similar calls and weights are localized making it easier to see the local relationships that two cats roar, two growl, and two purr; and the global relationship that the heavier the cat the more vocal its call. Modifying the natural category order only affects global relationship learning which depends on viewing multiple categories simultaneously. Notice it is easier to see the global relationship between call and weight in Matrix 1 where the categories call and weight are adjacent (and better localized) than in Matrix 2 where the categories are separated. In summary, good topical organization facilitates learning facts, local relationships, and global relationships, whereas good categorical localization primarily facilitates learning global relationships.

This research also revealed what does not work when learning from displays. Results consistently showed that learning from standard text or from outlines is less effective than learning from a matrix. These findings are consistent with previous research (see Kiewra 1994, for a review). A standard text fails to signal information, extract it, and localize it as does a matrix. An outline does all these things but its linear structure separates the related information across topics making it a less effective means for localization. Surprisingly, adding signals to the text or extracting its most important information did nothing to improve text learning. These new findings suggest that text learning requires an integration of ideas not made possible by simply signaling or extracting key ideas (Kiewra 2009). Those key ideas are best organized in a matrix that allows students to see relationships. Not any matrix will do, however. A matrix must be organized so that related facts are localized and relationships are readily apparent.

Although our findings appear valid, they are somewhat constrained. The case for validity is bolstered because learning materials were lengthy and realistic (with the exception of the extracted text that was used to achieve experimental control), tests assessed three types of learning outcomes, and because testing occurred on three occasions. The constraints are that the study materials were teacher generated and easily adapted to cross-classification. The first constraint means that although teachers or textbook authors can successfully produce matrix displays for students, it is not known empirically whether students can generate them as successfully on their own. One source for teaching students to generate matrices is *Teaching to Learn* (Kiewra 2009). The second constraint means that matrices are only appropriate when comparing multiple topics across one or more categories such as when comparing the planets in terms of diameters and rotation speed or when comparing polygons in terms of number of sides and area formula. Although the comparative cross-classification structure is one of several knowledge structures named by Jonassen and colleagues (Jonassen et al. 1993), it is perhaps the most pervasive. Anytime a topic is studied such as cognitive theory or cumulous clouds, such topics are ordinarily studied relative to associated topics such as behavioral theory and nimbus clouds, respectively. Whenever two or more topics are explored, a matrix is most effective.

Appendix

Wildcat Matrices 1, 2, 3 and 4 from Experiment 2.

Matrix 1 Organized topics, organized categories

	Tiger	Lion	Jaguar	Leopard	Cheetah	Bobcat
Call	Roar	Roar	Growl	Growl	Purr	Purr
Weight	Heavy	Heavy	Moderate	Moderate	Light	Light
Life span	Long	Long	Medium	Medium	Short	Short
Habitat	Jungle	Plains	Jungle	Jungle	Plains	Forest
Social behavior	Solitary	Group	Solitary	Solitary	Group	Solitary
Range	Confined	Vast	Confined	Confined	Vast	Confined

Matrix 2 Organized topics, random categories

	Tiger	Lion	Jaguar	Leopard	Cheetah	Bobcat
Habitat	Jungle	Plains	Jungle	Jungle	Plains	Forest
Weight	Heavy	Heavy	Moderate	Moderate	Light	Light
Social behavior	Solitary	Group	Solitary	Solitary	Group	Solitary
Call	Roar	Roar	Growl	Growl	Purr	Purr
Range	Confined	Vast	Confined	Confined	Vast	Confined
Life span	Long	Long	Medium	Medium	Short	Short

Matrix 3 Random topics, organized categories

	Leopard	Cheetah	Tiger	Bobcat	Lion	Jaguar
Call	Growl	Purr	Roar	Purr	Roar	Growl
Weight	Moderate	Light	Heavy	Light	Heavy	Moderate
Life span	Medium	Short	Long	Short	Long	Medium
Habitat	Jungle	Plains	Jungle	Forest	Plains	Jungle
Social behavior	Solitary	Group	Solitary	Solitary	Group	Solitary
Range	Confined	Vast	Confined	Confined	Vast	Confined

Matrix 4 Random topics, random categories

	Leopard	Cheetah	Tiger	Bobcat	Lion	Jaguar
Habitat	Jungle	Plains	Jungle	Forest	Plains	Jungle
Weight	Moderate	Light	Heavy	Light	Heavy	Moderate
Social behavior	Solitary	Group	Solitary	Solitary	Group	Solitary
Call	Growl	Purr	Roar	Purr	Roar	Growl
Range	Confined	Vast	Confined	Confined	Vast	Confined
Life span	Medium	Short	Long	Short	Long	Medium

References

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*, 155–159. doi:10.1037/0033-2909.112.1.155.
Day, R. S. (1988). Alternative representations. In G. Bower (Ed.), *The Psychology of learning and motivation* (Vol. 22, pp. 261–303). New York: Academic Press.
Igo, L. B., & Kiewra, K. A. (2007). How do high-achieving students approach web-based copy and paste note taking? Selective pasting and related learning outcomes. *Journal of Advanced Academics, 18*, 512–529.
Igo, L. B., Kiewra, K. A., & Bruning, R. B. (2008). Individual differences and intervention flaws: A sequential explanatory study of college students' copy-and-paste note taking. *Journal of Mixed Methods Research*, *2*, 149–168
Jairam, D., & Kiewra, K. A. (2009). Helping students soar to success on computers: An investigation of the SOAR study method. Manuscript submitted for publication
Jairam, D., & Kiewra, K. A. (in press). An investigation of the SOAR study method. *Journal of Advanced Academics.*
Jonassen, D. H., Beissner, K., & Yacci, M. (1993). *Structural knowledge: Techniques for representing, conveying, and acquiring structural knowledge*. Hillside, NJ: Lawrence Earlbaum.
Kauffman, D. F. (1998). *Theoretical constructs and practical applications of graphic organizers: Is the matrix more effective than text with review questions*? Unpublished masters thesis, University of Nebraska, Lincoln, Nebraska.
Kauffman, D. (2004). Self-regulated learning in Web-based environments: Instructional tools designed to facilitate self-regulated learning. *Journal of Educational Computing Research, 30*, 139–162. doi:10.2190/AX2D-Y9VM-V7PX-0TAD.
Kauffman, D. F., & Kiewra, K. A. (1998). *The matrix organizer boosts relational learning more than text or outlines*. San Diego, CA: American Educational Research Association. April.
Kauffman, D. F., Zhang, G., & Yang, A. (2004). *Web-based note taking and metacognitive prompts: A self-regulated learning perspective*. Paper presented at the annual meeting of the American Educational Research Association, San Diego, CA.
Kiewra, K. A. (1994). The matrix representation system: Orientation, research, theory, and application. In J. Smart (Ed.), *Higher education: Handbook of theory and research*. New York: Agathon.
Kiewra, K. A. (2009). *Teaching how to learn: The teacher's guide to student success*. Thousand Oaks, CA: Corwin Press.
Kiewra, K. A., Dennison, R. S., & Benton, S. T. (1995). *How studying text supplements affects prose processing*. Paper presented at American Educational Research Association, San Francisco, CA (April).
Kiewra, K. A., Dubois, N. F., Christian, D., & McShane, A. (1988). Providing study notes: Comparison of three types of notes for review. *Journal of Educational Psychology, 80*, 595–597. doi:10.1037/0022-0663.80.4.595.
Kiewra, K. A., Dubois, N. F., Christian, D., McShane, A., Meyerhoffer, M., & Roskelley, D. (1991). Note-taking functions and techniques. *Journal of Educational Psychology, 83*, 240–245. doi:10.1037/0022-0663.83.2.240.
Kiewra, K. A., Kauffman, D. F., Robinson, D. F., Dubois, N. F., & Staley, R. K. (1999). Supplementing floundering text with adjunct displays. *Instructional Science, 27*, 373–401.
Kiewra, K. A., Lang, J., Phifer, M., & Schraw, G. (1995b). *How graphic organizers affect learning from a chapter-length passage*. San Francisco, CA: American Educational Research Association.

Kiewra, K. A., Schraw, G., Lang, J., & Phifer, M. (1994). *How graphic organizers affect text learning: Factual, relational, and transfer effects*. Paper presented at the annual meeting of the Midwestern Educational Research Association, Chicago, IL.

Larkin, J. H., & Simon, H. A. (1987). Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science, 11*, 65–99. doi:10.1016/S0364-0213(87)80026-5.

Lorch, R. F. (1989). Text-signaling devices and their effects on reading memory processes. *Educational Psychology Review, 1*(3), 209–234. doi:10.1007/BF01320135.

Mayer, R. E. (2002). *The promise of educational psychology*. Upper Saddle River, NJ: Merrill Prentice Hall.

McCrudden, M., Schraw, G., Hartley, K., & Kiewra, K. A. (2004). The influence of presentation, organization, and example context on text learning. *Journal of Experimental Education, 72*, 289–306. doi:10.3200/JEXE.72.4.289-306.

Meyer, B. J. F., Brandt, D. H., & 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. doi:10.2307/747349.

Moreno, R., & Mayer, R. E. (2002). Verbal redundancy in multimedia learning: When reading helps listening. *Journal of Educational Psychology, 94*, 156–163. doi:10.1037/0022-0663.94.1.156.

Robinson, D., Katayama, A., DuBois, N., & DeVaney, T. (1998). Interactive effects of graphic organizers and delayed review on concept acquisition. *Journal of Experimental Education, 67*, 17–31.

Robinson, D. H., Katayama, A. D., Odom, S., Beth, A., Hsieh, Y. P., & Vanderveen, A. (2006). Increasing text comprehension and graphic note-taking using a partial graphic organizer task. *The Journal of Educational Research, 100*, 103–111. doi:10.3200/JOER.100.2.103-111.

Robinson, D. H., & Kiewra, K. A. (1995). Visual argument: Graphic organizers are superior to outlines in improving learning from text. *Journal of Educational Psychology, 87*, 455–467. doi:10.1037/0022-0663.87.3.455.

Robinson, D. H., & Schraw, G. (1994). Computational efficiency through visual argument: Do graphic organizers communicate relations in text too effectively? *Contemporary Educational Psychology, 91*, 399–415. doi:10.1006/ceps.1994.1029.

Robinson, D. H., & Skinner, C. H. (1996). Why do graphic organizers facilitate search processes: Fewer words or computational efficiency? *Contemporary Educational Psychology, 21*, 166–180. doi:10.1006/ceps.1996.0014.

Schmidt, S. (1991). Can we have a distinctive theory of memory? *Memory & Cognition, 19*, 523–542.

Titsworth, S., & Kiewra, K. A. (2004). Organizational lecture cues and student notetaking. *Contemporary Educational Psychology, 29*, 447–461. doi:10.1016/j.cedpsych.2003.12.001.

Van Djik, T. A., & Kintch, W. (1983). *Strategies of discourse comprehension*. New York: Academic Press.

Winn, W. (1991). Learning from maps and diagrams. *Educational Psychology Review, 3*, 211–247. doi:10.1007/BF01320077.