APPLIED COGNITIVE PSYCHOLOGY
Appl. Cognit. Psychol. **20**: 917–940 (2006)
Published online 19 June 2006 in Wiley InterScience
(www.interscience.wiley.com). DOI: 10.1002/acp.1236

Learning Procedures: The Role of Working Memory in Multimedia Learning Experiences

TAD T. BRUNYÉ[1]*, HOLLY A. TAYLOR[1], DAVID N. RAPP[2] and ALEXANDER B. SPIRO[3]

[1]*Department of Psychology, Tufts University, USA*
[2]*Department of Educational Psychology, University of Minnesota, USA*
[3]*Department of Psychology, Tufts University and McLean Hospital, USA*

SUMMARY

The ubiquitous label 'some assembly required' signals the appearance of instructions for assembly procedures. These instructions come in various formats, some of which may be more effective than others. Previous research has demonstrated advantages for multimedia as compared to single-format presentations. The current study sought to outline the cognitive processes contributing to this advantage. Specifically, two experiments examined the working memory and source monitoring processes involved with remembering procedural instructions presented in three different formats. Participants learned procedural instructions while undertaking one of a variety of selective interference tasks targeting working memory subcomponents. Results, while supporting a multimedia advantage for learning, demonstrated selective working memory subsystem involvement with different instruction formats. Further, despite the multimedia advantage, participants often misremembered multimedia presentations as picture-based ones. These results provide further insight into the cognitive processes that underlie comprehension and memory for multimedia experiences. Copyright © 2006 John Wiley & Sons, Ltd.

Whether rebuilding a car engine, folding a paper airplane, assembling a stereo surround sound system, or following a recipe, the overall goal is similar: to follow a series of steps through to a final product. Instruction manuals (whether paper-based, CD- or DVD-ROM) and human tutors (e.g., teachers in a car repair or cooking course) rely on a variety of methods for detailing such procedures. Their goal, of course, is to help the learner not only construct a completed product, but also to learn the relations between procedural steps. Building such familiarity leads to facile use of the knowledge so that the procedures can be completed in different situations (e.g., outside of the classroom or without the manual). For example, a task might necessitate reversing well-learned steps, so as to not only put an object together, but also take it apart (e.g., systematically dismantling a car engine to diagnose mechanical problems; breaking down a completed recipe into components to improve it). Thus, an important issue in the study of procedural learning involves determining effective methods for helping individuals build strong, manipulable representations of task steps that will lead to reproducible, reliable outcomes.

*Correspondence to: T. T. Brunyé, Department of Psychology, Tufts University, 490 Boston Ave., Medford, MA 02155, USA. E-mail: tbruny01@tufts.edu

Multimedia has rapidly become a popular format for presenting such information. Multimedia includes any presentation combining more than one format (e.g., words and pictures; Mayer, 2001), whether it be within a single sensory modality (e.g., a visual display of steps to construct a prefabricated table along with printed text) or across modalities (e.g., a video with voice-over narration on pet training). Its use in educational and applied contexts has become commonplace; specifically combinations of pictures and text now appear in a variety of mediums, including books, print and electronic advertisements, websites, and animation-based assembly instructions delivered through computer software.

Research exploring knowledge acquisition from multi-format sources has largely supported the effectiveness of multimedia relative to single-format learning, particularly for expository, fact-based content such as newspapers and science explanations (see Allen, 1971; Kozma, 1991; Levie & Lentz, 1982; Mayer, 2001; and Peeck, 1994 for reviews). However, expository content is only one type of discourse genre within which people may learn (Zwaan, 1994). Another form, the type we have outlined above, involves procedural information. In fact, the ubiquitous phrase 'some assembly required' is a cue to prepare to encode procedural information. Procedures, unlike expository information, place particular, often explicit emphasis on temporal order and spatial relationships.

Multimedia, specifically text and pictures, may in fact facilitate comprehension of procedural information by explicitly conveying the temporal and spatial nature of assembly instructions. Such multimedia instructions could, for example, employ text to describe sequential information, while using pictures to illustrate spatial relations between components. Additionally, multimedia benefits may accrue not just from such explicit cues, but also due to the nature of human information processing (e.g., visual and spatial subsystems in working memory, Baddeley, 1992). Only a handful of studies have investigated the underlying cognitive processes involved in learning procedural information through multimedia presentations (e.g., Diehl & Mills, 1995; Glenberg & Langston, 1992; Marcus, Cooper, & Sweller, 1996; Novick & Morse, 2000; Stone & Glock, 1981; Zacks & Tversky, 2003). Even fewer studies have investigated the limitations of multimedia as a function of these processes (Marcus et al., 1996). The present research was designed to further explicate the cognitive mechanisms underlying potential multimedia advantages, as well as their limits, by focusing on presentation format differences in the relatively understudied genre of procedural learning. To do this, we utilized several dependent measures for assessing the representations learners build as a function of these presentations. We also assessed the degree to which such representations may fall victim to human processing limitations (e.g., working memory processes and source monitoring difficulties). The implications derived from these investigations help delineate the cognitive mechanisms underlying procedural learning from multimedia experiences, and inform design guidelines for procedural multimedia applications.

IS THERE A MULTIMEDIA ADVANTAGE?

Research concerning multimedia utility has tended to focus on the inclusion of pictures and diagrams with expository texts (e.g., Glenberg & Langston, 1992; Gyselinck & Tardieu, 1999; Gyselinck, Cornoldi, DuBois, De Beni, & Ehrlich, 2002; Mayer, 1989; Mayer & Gallini, 1990; Mayer, Bove, Bryman, Mars, & Tapangco, 1995; Peeck & Jans, 1987; Stone & Glock, 1981; also see Kozma, 1991; and Levie & Lentz, 1982 for reviews). Some of these studies have focused on particular combinations of multimedia, such as combining narration and animation with pictures and text, respectively (Baggett, 1979; Mayer & Anderson, 1991; Mayer & Moreno, 2002). Of course, the most basic form of multimedia combines text and pictures in a single presentation. To assess multimedia effectiveness, we examined this basic situation, comparing picture-text combinations to single-format text and single-format picture presentations. We also selected this multimedia combination based on growing evidence of its effectiveness in facilitating comprehension of expository information. This has been demonstrated as a function of performance on immediate and delayed recall (e.g., Mayer, 1989; Mayer & Gallini, 1990; Mayer et al., 1996; Peeck & Jans, 1987), recognition (e.g., Mayer, 1989), transfer problems (e.g., Mayer, 1989; Mayer & Gallini, 1990; Mayer et al., 1996), sequential order verification (Glenberg & Langston, 1992), and object assembly (Novick & Morse, 2000).

There are, as well, limitations to multimedia's effectiveness, which emphasizes the importance of careful design and implementation of format and modality combinations (Novick & Morse, 2000; also see Mayer & Moreno, 2002 for a review). With this in mind, and to serve as an appropriate framework to our study, we next provide a review of multimedia research, emphasizing multi-format (e.g., pictures/diagrams and text) rather than multi-modal (e.g., visual, auditory) presentations. Following that discussion, we consider the underlying processes that may influence multimedia comprehension, with particular attention to working memory. We conclude our introduction with a focus on the specific role multimedia may play in remembering procedural information.

RESEARCH ON MULTIMEDIA EXPERIENCES

Levie and Lentz (1982) comprehensively review published studies, up to that date, showing the relative efficacy of texts with and without illustrations (a review of early instructional media research is beyond the scope of this article; see Allen, 1967, 1971 for discussion of historical context). This review categorized studies into three multimedia groups that were compared with text-only presentations: (1) text with relevant pictures; (2) text with irrelevant pictures; and (3) text with vaguely-defined relevant and irrelevant pictures, not otherwise specified. The review reported that relevant illustrations were beneficial to text learning in the majority (85%) of comparisons, and no study indicated a relative advantage for text alone. In contrast, irrelevantly-illustrated text fared no better than text alone in the vast majority (90%) of reviewed studies. The final comparison, involving vaguely-defined pictures with text, revealed mixed findings, with either advantages for those texts or no differences as compared to text alone. Overall, early multimedia research supported the view that adding pictures to text could enhance learning, with the caveat that illustration relevance plays an important role.

These studies examined situations involving the addition of illustrations to stand-alone texts; however, we might also consider cases wherein text is added to stand-alone illustrations (such as captions accompanying figures in scientific articles). Instructional diagrams without text are commonly found in object assembly instructions (e.g., furniture and do-it-yourself craft guides), particularly when those instructions are intended for multiple countries with different language preferences. While the above studies looked at texts alone versus texts with pictures, relatively few studies (e.g., Gyselinck et al., 2002; Stone & Glock, 1981) have compared pictures alone to picture/text combinations. For a valid comparison, the two formats would need to be informationally equivalent; in other words, a picture would have to convey roughly the same informational content as the corresponding text. Indeed, some of the studies assessing pictures alone versus multimedia have not equated informational content. In the case of Gyselinck et al. (2002), their 2-dimensional pictures provided complementary rather than equivalent information, and the objects in those pictures were not easily identifiable. For example, in conveying static electricity conductance and transfer concepts, the authors used explanatory text for identifying items (e.g., ebonite rod, fur, and a plexiglass ruler) and understanding their complex relationships (i.e., becoming charged with positive or negative electricity). Perhaps not surprisingly in this case, pictures alone could not support the same inference and paraphrase performance as multimedia. Students would be unlikely to identify such relationships from the pictures without additional information.

Object assembly instructions may provide a domain for which a picture-only group is not only feasible, but also might actually be expected and commonplace. In line with this view, Stone & Glock (1981) asked participants to study procedures presented as either text-only, 3-dimensional pictures-only, or text with accompanying pictures. As they studied these procedures, participants also assembled the objects. Picture-only and multimedia instructions resulted in fewer assembly errors than text-only instructions. Thus, the authors concluded that 'illustrations convey spatial information more effectively than text' (p. 425), congruent with the notion that pictures should facilitate memory for procedures because they directly illustrate spatial relations (Garrison, 1978; Glenberg & Langston, 1992; Levie & Lentz, 1982). Indeed, for procedural instructions, pictures can directly depict (rather than indirectly describe) the actions necessary to complete a construction task (e.g., Zacks & Tversky, 2003).

In summary, the majority of studies examining multimedia efficacy suggest substantial benefits from adding relevant illustrations to text descriptions. Fewer studies have made analogous comparisons by adding text to pictures. One potential reason for this imbalance may be the previous focus on expository information; constructing picture-only presentations for expository information is difficult, as expository presentations often necessitate explicit description and extended explanation. In contrast, procedures (as demonstrated by Stone & Glock, 1981) can provide a domain for which informationally similar picture-only and text-only instructions are both easily constructed and often expected. What this work has not assessed is the degree to which the addition of text to individually depicted procedural steps might aid memory (as opposed to online assembly performance) for such procedures. Stone and Glock's (1981) reliance on a single measure (i.e., assembly) did not obtain differential results for picture-only and multimedia displays. This suggests, counter to the previously described literature, that multimedia may only result in an advantage in comparison to text, at least for object assembly. Thus, one concern may be the degree to which multimedia advantages are truly a function of the interactions between text and pictures, or purely a function of picture-based benefits. We return to this issue in our experiments; for now, we discuss the underlying processes that may impact benefits attributed to multimedia presentations.

UNDERLYING MECHANISMS OF MULTIMEDIA LEARNING

Mayer (1997) proposed the *Generative Theory of Multimedia Learning* to provide a cognitive account of the advantages (and limitations) of multimedia presentations. The theory pulls together classic findings on memory functioning, including dual coding theory

(Paivio, 1986) and the generation effect (Soraci et al., 1994), to account for a wide range of outcomes from multimedia studies. We next discuss these classic theories and the role they play in Mayer's theory.

Dual-coding theory posits simultaneous and independent processing of a stimuli's verbal and image components (e.g., Paivio, 1986). According to this view, information can be encoded in two formats: visual and verbal. Information encoded in two forms is more readily retrievable from memory compared to information in a single format. Classic evidence supporting this finding comes from work on memory for nouns: concrete nouns (e.g., cat) are more easily remembered than abstract nouns (e.g., virtue), as concrete nouns can be coded in both visual and verbal formats, while abstract nouns tend to be coded in a purely verbal format (Paivio, 1965). The *Generation effect* proposes that active selection and integration within working memory facilitates memory (e.g., Wittrock, 1989). When individuals generate their own interpretations or items in memory tasks, they recall that information more accurately than when items are provided in a more passive manner (Slamecka & Graf, 1978). In general, the more active a participant is in selecting and integrating information, the more comprehensive the resulting mental representation. The ability to simultaneously process text and images, in conjunction with the active integration and formation of new memories, helps to create powerful memory associations. These relatively strong mental representations are theorized to be responsible for the multimedia advantage.

The combination of these two processing theories represents a substantial step towards understanding the cognition of multimedia learning (Mayer, 1997). Specifically, investigations of working memory (see Baddeley, 1992) during multimedia learning have demonstrated selective interference effects consistent with separable phonological and visuospatial processing (Gyselinck et al., 2002). Finger-tapping or syllable-string repetition secondary tasks selectively interfere with spatial (images) and verbal (text) processing, respectively. The visuospatial secondary task reduces the advantage of adding pictures to text; conversely, the articulatory secondary task selectively reduces text effectiveness. Pictures and text within multimedia presentations, thus, may be processed by individual working memory subsystems prior to integration. However, the Generative Theory of Multimedia Learning (Mayer, 1997) also hypothesizes integration while the spatial and verbal materials are active within working memory. The allocation of these subsystem resources should occur via the central executive (Baddeley, 1992, 1996; Logan, 1985; Smith & Jonides, 1999). While multimedia theories have focused on subsystem activity, they have tended to ignore the potential role of the central executive. Defining the impact of central executive activity is critical for a complete understanding of the cognitive mechanisms underlying multimedia processing. The present studies, therefore, examine whether secondary central executive tasks selectively interfere with the allocation of resources and successful integration of text and pictures (as is the case with multimedia), as predicted by Mayer's (1997) and Baddeley's (1992) models. Investigating these subsystems provides a test of the appropriateness of the dual-coding and generative theories for explaining multimedia benefits.

PROCEDURAL APPLICATION AND PRESENT HYPOTHESES

As described earlier, procedural information provides a genre wherein text-only, picture-only, and multimedia presentations commonly occur. Additionally, unlike expository material, multimedia procedural presentations rely on redundancy across pictures and text, and thus the relative effectiveness of texts and pictures can be assessed while reducing potential confounds due to different informational content. In other words, texts and pictures need not serve a purely complementary role as is commonly the case in expository learning experiences. However, procedural presentations are not without difficulties; other differences must be considered based on the extant literature on expository multimedia. Specifically, procedural information requires sequential encoding to maintain the separation and cumulative interdependence of steps, information that may not be required for a mental representation of expository information. In addition, the spatially demanding nature of procedural sequences may uniquely benefit from picture-only presentations, which can simultaneously convey object identities, shapes, orientations and spatial relationships. Thus, while procedural learning may be amenable to assessing the potential benefits of multimedia experiences, there are several mediating factors that may affect the generalizability of such benefits to other genres.

Nevertheless, with regard to Mayer's (1997) theory, working memory processes involved in procedural multimedia should be the same as those involved in expository stimuli. While single-format procedural presentations are processed by individual working memory subsystems (i.e., pictures by the visuospatial component; text by the phonological component), dual-format presentations should recruit both subsystems. Descriptions conveying visuospatial information, however, may be a special case in that they appear to recruit both phonological and visuospatial mechanisms, as evidenced by recent work investigating the representation of spatial descriptions (e.g., De Beni, Pazzaglia, Gyselinck, & Meneghetti, 2005). Each presentation type may also differentially rely on central executive resources for integration between the two subsystems, with multimedia relying on integration to a greater extent (e.g., Baddeley, 1992). The resulting mental representations based on dual coding should lead to robust advantages relative to single-format learning. The active selection (e.g., Wittrock, 1989) of presentation elements (labels, parts, relationships etc.) should additionally lead to learning advantages when participants have more than one format from which to gather information. In accordance with previous procedural multimedia research (Stone & Glock, 1981), pictures in spatially-demanding contexts should be particularly effective if they are informationally similar to their text counterparts. In the case of multimedia, participants can be expected to rely primarily upon the spatially-rich images for content, and perhaps rely on the text to resolve labeling or associative ambiguities. Therefore, pictures are expected to produce better memory performance relative to text, and multimedia performance should exceed both picture-only and text-only performance. However, to the extent that procedural presentations uniquely promote reliance on spatially-rich images, or individuals are biased to study pictures over text, participants may demonstrate attentional biases towards pictures. This should result in equivalent memory performance between multimedia and picture-only conditions.

If, as might be suggested, procedural multimedia is processed by working memory in a manner analogous to other discourse genres such as expository multimedia, secondary tasks should selectively reduce the advantages of targeted formats. This prediction suggests invariant selective interference across content genre as a function of subsystem activation. Specifically, an articulatory secondary task should selectively interfere with verbal processing (see Farmer, Berman, & Fletcher, 1986), leading to performance reductions in the text-only condition, and more reliance on pictures in the multimedia condition. Conversely, a visuospatial secondary task should selectively interfere with picture processing (see Kruley, Sciama, & Glenberg, 1994; Logie, 1995), leading to performance decrements following picture-only presentations, and if our hypothesis on the reliance on pictures within multimedia is correct, on multimedia learning as well. Beyond these two subsystems, we can also consider secondary tasks as a means for assessing central executive processes in multimedia learning. Although the effects of a secondary task designed to recruit central executive resources has not been examined in a multimedia context, one could expect interference from such a secondary task primarily in multimedia processing, when dual-format integration is explicitly required (Mayer, 1997). Past research demonstrates central executive recruitment in tasks requiring random generation of either verbal digits or manual key presses (Baddeley, 1996; Baddeley, Emslie, Kolodny, & Duncan, 1998); thus, such tasks might interfere with general central executive processing. Additionally, secondary central tasks that recruit articulatory or visuospatial resources may interfere with their respective working memory subsystems.

The present experiments examine working memory processes during the comprehension of procedural information as a function of presentation format. The inherent sequential nature of procedural information demands appropriately targeted dependent measures. The present studies investigate memory for those procedures through recall and sequential order verification, using a task similar to that of Glenberg and Langston (1992). We assessed accuracy and latencies to order verification judgments as a means of examining format effects on procedural learning. Additionally, we examined participants' phenomenological experiences with the presentations, particularly with respect to memory for presentation format. To do this, we used a source monitoring task to examine participants' ability to recall the original learning format (e.g., Johnson, Hashtroudi, & Lindsay, 1993). These combined dependent measures afford a more comprehensive analysis of any multimedia advantage, and help outline the roles of underlying memory processes during multimedia examples of procedural presentations.

EXPERIMENT 1

We began our experiments by assessing not only the impact of different types of presentations on procedural learning, but also the selective interference of secondary articulatory tasks (assumed to interfere with concurrent verbal activities). In Experiment 1, participants learned procedural assembly sequences, each in one of three formats: picture, text, or multimedia (the combination of pictures and text). Participants were randomly assigned to one of five secondary task groups:

1. A control group with no secondary task.
2. An articulatory task involving the repetition of learned syllable strings.
3. A visuospatial task involving the repetition of learned finger tap sequences.
4. A central-articulatory task requiring random generation of syllable strings.
5. A central-visuospatial task requiring random generation of finger taps.

Our dependent measures included order verification accuracy and latency, free recall, and format recall.

Method

Participants

Seventy-five Tufts University undergraduates participated for partial course credit.

Materials

Assembly instructions. Eighteen five-step pictorial assembly sequences from Kinder Egg™ toys were selected using the following constraints: each sequence comprised five steps, with the last step depicting the finished product; sequences had easily identifiable parts; no two sequences were similar to each other with respect to parts or spatial organization; all sequences were in color format. From these pictorial instructions we asked seventeen pilot participants to generate sequence titles and a descriptive text for each pictorial sequence step. These descriptions were then collapsed into one coherent and concise sequence title and statement for each sequence step. This task was designed to create text statements that were roughly equivalent to, not complementary to, the picture-based instructions. Pictorial stimuli were 300 × 300 pixels in size and text stimuli were presented in bold 14-point Times New Roman font. Multimedia instructions combined the pictorial and text instructions on one screen, with pictures above text. Picture-only and text-only instructions were presented twice in the same format, with the same picture or text being presented simultaneously on a single screen (see Figure 1 for examples of these presentations). This resulted in a total of 54 assembly sequences (eighteen each of picture-only, text-only, and multimedia).

Secondary tasks. Four self-paced secondary tasks were used, two verbal (one articulatory and one central-articulatory) and two spatial (one visuospatial and one central-visuospatial). The articulatory task was a syllable string repetition task (adapted from Gyselinck et al., 2002) that required participants to repeat the sequence 'BA BE BI BO.' The central-articulatory task required participants to produce the syllables 'BA BE BI BO' in a non-sequential order (i.e., in a random-like fashion, continuously). Thus, while both tasks require articulatory processes (syllable verbalization), the central-articulatory task also necessitates decisions about which syllable to select and produce. Such random production tasks have been shown to recruit central executive resources (e.g., Baddeley, 1996; 1998).

Two finger tapping tasks were developed, similar to those used in Gyselinck et al. (2002). The visuospatial secondary task required participants to tap four keys (2, 4, 6, 8) on a number keypad in a set counter-clockwise rotation at a rate approximating 1 tap per second. The keys were arranged in a cross, and were hidden from view. Finger tapping was

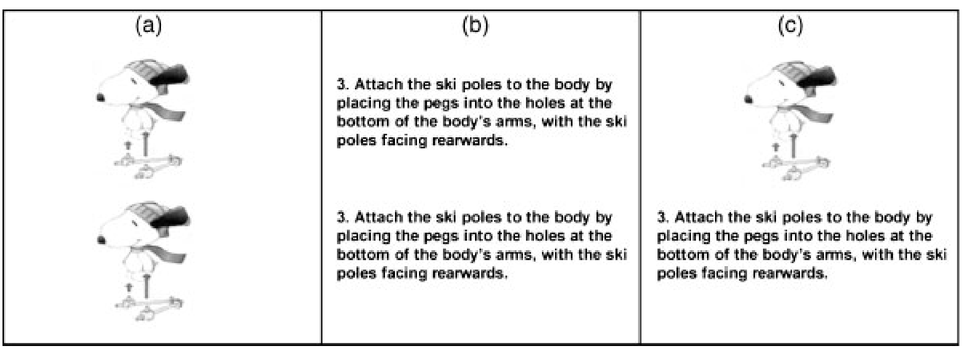

Figure 1. a–c. Picture only (a), text only (b), and multimedia (c) versions of a single step in a given assembly sequence

done using a participant's dominant writing hand. The central-visuospatial task required participants to randomly produce sequences of finger taps on the same keypad.

Dependent measures. The present experiment incorporated three memory tasks: order verification, free recall, and format recall.

The order verification task was adapted from Glenberg and Langston (1992). This task required participants to verify whether two sequence steps, presented from left to right, appeared in the correct temporal order, even if that order did not involve sequential steps. For example, when presented step 2 (on the left) and step 4 (on the right), participants should report that these steps appeared in this order, even though step 3 occurred between them. In contrast, if step 4 appeared on the left and step 2 on the right, participants should respond that the steps did not appear in this order. The task included 180 comparisons, 10 for each of the 18 assembly sequences, divided into three equal blocks. Half of the trials depicted the correct temporal order and half an incorrect, reversed order. Half of the trials presented steps in pictures and half in text, regardless of the original learning format. For the picture comparisons, the images were modified from their original versions to prevent participants from responding based simply on how 'complete' the picture appeared (i.e., a picture that appeared more complete should always appear to the right of a less complete picture). In these modified pictures, the critical component of each step (e.g., broom) was applied to a fully assembled toy (e.g., snowman). Additionally, the trials included four 1-step comparisons (e.g., step 2 compared to step 3), four 2-step comparisons (e.g., step 1 compared to step 3), and two 3-step comparisons (e.g., step 1 compared to step 4).

For the free recall task participants received a numbered sheet and were asked to write in the title and the five steps of each sequence. Participants were asked to recall the steps of a procedure in the correct order to the best of their ability, but to recall sequences in any order if they were unsure of the exact ordering.

For the format recall task, participants received a sheet of paper listing the 18 sequence titles, and were asked to circle a label designating the format in which they had learned a sequence (i.e., picture-only, text-only, or pictures and text).

Procedure

Learning. The 18 assembly sequences were presented in three blocks of six. Within each block, two sequences were picture-only, two were text-only, and two were multimedia. Sequences within each block were randomly presented. Counterbalancing ensured that different participants received the same sequence in different formats. Our multimedia presentations were inherently redundant since the picture and text provided similar information, as a function of the pilot-produced descriptions. To control for this conceptual redundancy (i.e., picture above text), picture-only and text-only presentations included the same information twice (picture above picture, or text above text) simultaneously. Although we could not guarantee that participants would read both statements (or view both pictures), we also understand that they could elect to ignore one of the two multimedia formats (we will return to this issue in the General Discussion). Sequence titles (e.g., 'Assembling a Snoopy on skis') appeared for 5 seconds immediately preceding each sequence, and sequence steps were presented at a rate of 10 seconds each.

Secondary tasks. Participants were divided into five secondary task groups: control (no secondary task), articulatory, central-articulatory, visuospatial, and central-visuospatial.

Participants received instructions for and practiced their assigned secondary task until the experimenter noted perfect performance. Specifically, for the central-articulatory and central-visuospatial tasks, the experimenter recognized any perceptible articulation or tapping patterns and verbally encouraged randomness. All secondary tasks were self-paced, and approximated one response (tap, or syllable) per second.

Memory tasks. After each presentation block, participants completed a 7×7 grid maze to clear working memory. They then completed the free recall followed by the order verification task. This ordering, based on previous pilot research, minimizes carry-over effects introduced by the order verification task (Brunyé, Taylor, & Rapp, 2003). Specifically, participants have been found to relearn sequences (content and order) during the order verification task and apply that information at recall. However, free recall preceding order verification did not affect order verification performance. During free recall, participants recalled as many and as much of the six sequences in the preceding block as possible, in a self-paced manner. After completing the free recall, participants began the order verification task. Because pictures required modification for this task, participants received instructions about these changes and completed a self-paced practice session. During practice participants saw a novel 3-step sequence followed by two practice trials. The order verification task required participants to press 'YES' or 'NO' keys (C and M on the keyboard, respectively) to indicate whether the two steps were in the correct order. Dependent measures included response time (RT) and accuracy. Participants were asked to complete this task as quickly as possible without compromising accuracy. Finally, after all three learning-test blocks, participants completed the format recall task.

Results

Scoring

Free recall. Recall performance was scored by calculating the number of correct steps recalled, the number of titles recalled, and the number of step order errors. Accurate recall of an individual step required correct object labels (e.g., wheels, axle) and spatial relationship descriptions (*attach* wheels *to the ends of* axle). Labeling difficulties due to specialized vocabulary were counted as correct when participants made an effort to describe the object (e.g., the *axle* could also be described as the *long black rod*; the *bow of a boat* could also be called the *front of a boat*). Due to extremely low free recall error rates for both step order and title recall, only step recall accuracy data was analyzed.

Format recall. Recall performance was scored by counting and classifying the number of errors. Classification of errors involved designating the format of the presentation during study as compared to participants' recall of that format. For example, if a participant learned a picture-only sequence, but identified it as having been learned as multimedia (picture and text), this was classified as a picture-multimedia error (P-MM). This classification resulted in six potential error types (P-T, P-MM, T-P, T-MM, MM-T, MM-P).

Analyses

Two sets of analyses were run to examine the main issues of interest: 1) the effect of learning format and 2) the effect of different secondary on learning format. To investigate the first issue, the relative effectiveness of each learning format, we ran analyses on the control group data. Using the control group data alone gives a purer look at learning format effects since it is free from secondary task effects. The analysis for the free recall and order verification tasks consisted of a (learning condition: picture, text, multimedia) repeated measures ANOVA followed up with planned comparisons. The planned comparisons explicitly examined differences between multimedia and single-format (picture-only/text-only) conditions, and between picture-only and text-only conditions. To examine the effect of learning format on format recall we used a factorial (error type: P-T, P-MM, T-P, T-MM, MM-T, MM-P) repeated measures ANOVA for each learning condition.

To investigate the second issue, we first examined overall effects with an omnibus ANOVA for free recall, order verification, and format recall data. Effects of secondary task were then examined separately for each learning format. This involved three (picture, text, multimedia) one-way ANOVAs for the free recall and order verification data. We followed up on significant secondary task effects using planned comparisons between the control and two secondary groups (articulatory and central-articulatory). To examine for between-groups differences of format recall error rates, six one-way repeated measures ANOVAs were conducted on data from the two secondary task groups relative to the control group, and planned comparisons further probed between-group differences.

Format advantages

See Table 1 for control group means for the free recall and order verification tasks.

Free recall. The analysis of control group free recall performance revealed an effect of learning format, $F(2, 28) = 15.366$, $p < 0.01$, $MSE = 0.032$ (see Figure 2a). Planned comparisons revealed better recall following multimedia learning relative to picture-only ($p < 0.05$, Mean Difference = 0.203) and text-only ($p < 0.01$, Mean Difference = 0.364) formats, and better recall following picture-only compared to the text-only ($p < 0.05$, Mean Difference = 0.161) format.

Order verification. Analysis of control group order verification accuracy revealed a learning format effect, $F(2, 28) = 13.371$, $p < 0.01$, $MSE = 0.009$ (see Figure 2b). Planned comparisons revealed higher accuracy following multimedia relative to both picture-only

Table 1. Experiments 1 and 2 control group means and standard deviations for order verification (O.V.) task accuracy and reaction time, as well as instructions recall task accuracy as a function of presentation type

	Presentation type					
	Picture		Text		Multimedia	
Experiment and measure	*M*	*SD*	*M*	*SD*	*M*	*SD*
Experiment 1						
Free recall accuracy	0.567	0.259	0.406	0.231	0.769	0.116
O.V. accuracy	0.842	0.114	0.747	0.153	0.934	0.058
O.V. response time	6.84	2.03	6.97	2.82	5.09	1.87
Experiment 2						
Free recall accuracy	0.576	0.252	0.390	0.261	0.773	0.193
O.V. accuracy	0.843	0.135	0.738	0.152	0.923	0.074
O.V. response time	6.45	0.942	6.67	1.20	4.82	1.16

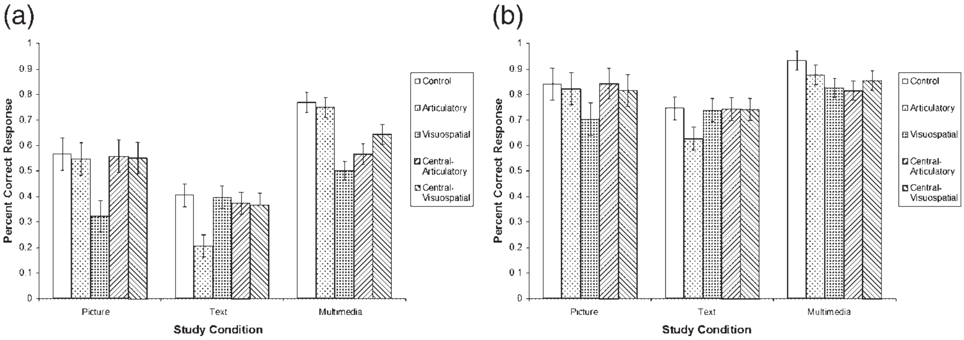

Figure 2. a–b. Experiment 1 mean control, articulatory, visuospatial, and the two central secondary task group accuracy rates and standard error derived from the free recall (a) and order verification (b) tasks for picture, text, and multimedia conditions

($p < 0.05$, Mean Difference = 0.092), and text-only ($p < 0.01$, Mean Difference = 0.186) formats. Participants also responded more accurately following picture-only compared to the text-only ($p < 0.05$, Mean Difference = 0.094) format.

Analysis of control group order verification RT (in seconds) revealed an effect of learning format, $F(2, 28) = 6.461$, $p < 0.01$, $MSE = 2.542$. Planned comparisons revealed faster RTs following multimedia presentations relative to picture-only ($p < 0.05$, Mean Difference = −1.743), and text-only ($p < 0.05$, Mean Difference = −1.875) formats. RTs following picture-only and text-only formats did not differ (p = n.s., Mean Difference = −0.132).

Format recall. Analysis of control group format recall error rates revealed an effect of error type, $F(5, 70) = 18.189$, $p < 0.01$, $MSE = 0.023$. Planned comparisons revealed higher MM-P error rates relative to all other types (all $ps < 0.01$); T-P ($M = 0.067$, $SD = 0.123$), T-MM ($M = 0.111$, $SD = 0.121$), P-T ($M = 0.033$, $SD = 0.069$), P-MM ($M = 0.089$, $SD = 0.106$), MM-T ($M = 0.033$, $SD = 0.069$), and MM-P ($M = 0.389$, $SD = 0.206$). See Figure 3 for a depiction of all six error rates.

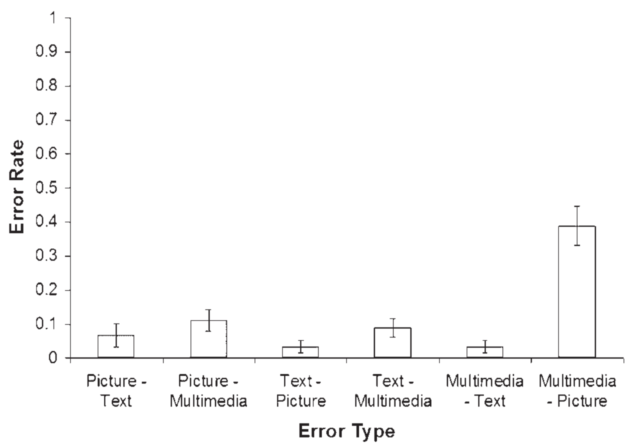

Figure 3. Experiment 1 mean occurrence and standard error for the six format recall test error type, for the control group

Selective interference from secondary tasks

Free recall. Figure 2a depicts free recall performance. The analysis revealed two primary findings. First, a main effect of learning condition, $F(2, 140) = 59.822$, $p < 0.01$, $MSE = 0.027$, demonstrated a pattern of results in line with control group analyses above, with multimedia producing higher recall rates relative to both picture-only and text-only formats. Second, a significant interaction between learning format and secondary task was obtained, $F(8, 140) = 5.257$, $p < 0.01$, $MSE = 0.027$, demonstrating significantly different secondary task influences across learning formats.

A one-way (between: articulatory, central-articulatory, visuospatial, central-visuospatial, control) ANOVA on recall of picture format procedures revealed an effect of secondary task, $F(4, 74) = 3.105$, $p = 0.021$, $MSw = 0.0529$. Planned comparisons showed higher recall in the control group relative to the visuospatial, $t(70) = 2.911$, $p < 0.01$, but not relative to any other secondary group (all $p > 0.05$). The analysis of recall of text format procedures revealed an effect of secondary task, $F(4, 74) = 2.93$, $p = 0.027$, $MSw = 0.0345$. Planned comparisons showed higher recall in the control group relative to the articulatory, $t(70) = 2.949$, $p < 0.01$, but not relative to any other secondary group (all $p > 0.05$). The analysis of recall of multimedia format procedures also revealed an effect of secondary task, $F(4, 74) = 6.015$, $p = 0.001$, $MSw = 0.0334$. Planned comparisons showed higher recall in the control group relative to the visuospatial, $t(70) = 4.034$, $p < 0.01$, and the central-articulatory, $t(70) = 3.036$, $p < 0.01$, secondary groups, but not relative to any other secondary group. The control group demonstrated marginally higher recall relative to the central-visuospatial secondary group, $t(70) = 1.871$, $p < 0.10$.

Order verification. Figure 2b depicts the order verification results. The analysis revealed two primary results. First, a main effect of learning condition, $F(2, 140) = 40.568$, $p < 0.01$, $MSE = 0.009$, demonstrated a pattern of results in line with control group analyses above, with multimedia producing higher recall rates relative to both picture-only and text-only formats. Second, a significant interaction between learning format and secondary task was obtained, $F(8, 140) = 4.158$, $p < 0.01$, $MSE = 0.009$, demonstrating significantly different secondary task influences across learning formats.

A one-way (between: articulatory, central-articulatory, visuospatial, central-visuospatial, control) ANOVA on order verification accuracy of picture format procedures revealed an effect of secondary task, $F(4, 74) = 4.034$, $p = 0.005$, $MSw = 0.0124$. Planned comparisons showed higher verification performance in the control group relative to the visuospatial, $\mathrm{t}(70) = 3.396$, $p < 0.01$, but not relative to any other secondary group (all $p > 0.05$). The analysis of order verification accuracy of text format procedures revealed an effect of secondary task, $F(4, 74) = 2.523$, $p = 0.049$, $MSw = 0.0159$. Planned comparisons showed higher accuracy in the control group relative to the articulatory group, $t(70) = 2.612$, $p < 0.01$, but not relative to any other secondary group (all $p > 0.05$). The analysis of order verification accuracy based on multimedia presentation revealed an effect of secondary task, $F(4, 74) = 2.889$, $p < 0.05$, $MSw = 0.0116$. Planned comparisons revealed higher accuracy in the control group relative to the visuospatial, $t(70) = 2.724$, $p < 0.01$, central-visuospatial, $t(70) = 2.034$, $p < 0.05$, and central-articulatory, $t(70) = 3.02$, $p < 0.01$, groups, but not relative to the articulatory group ($p > 0.05$). Analyses of order verification RT did not show any effects.

Format recall. Control group format recall results revealed low error rates for all but the MM-P error type, guiding a subsequent focus on this error type to test our hypotheses. A one-way ANOVA of MM-P format recall errors revealed an effect of secondary task, $F(4, 74) = 7.518, p < 0.01, MSw = 0.045$ (see Figure 4). Comparisons showed significantly higher MM-P error rates in the articulatory relative to the control group, $t(70) = 2.713$, $p < 0.01$. Further, significantly lower MM-P error rates were found for visuospatial relative to the control group, $t(70) = -2.714$, $p < 0.01$. To gain insight into the mechanisms responsible for producing such high MM-P error rates, exploratory analyses of learning trials followed by MM-P misattribution errors were conducted on free recall and order verification data. Analyses of these learning trials revealed significantly higher free recall, $t(69) = 4.248$, $p < 0.01$, and order verification, $t(69) = 4.248$, $p < 0.01$, accuracy rates following multimedia ($M = 0.657$, $SD = 0.205$, and $M = 0.860$, $SD = .155$, respectively), relative to picture-only ($M = 0.526$, $SD = 0.239$, and $M = 0.807$, $SD = 0.122$, respectively) learning. These analyses have important implications for source monitoring in multimedia learning, and will be further interpreted in the Discussion section.

Discussion

Congruent with the Generative Theory of Multimedia Learning (Mayer, 1997), the results showed a multimedia advantage and extended this advantage to procedural materials. These latter materials necessitate knowledge about spatial relationships and sequential order. Participants recalled more information, and made faster and more accurate order verification judgments following multimedia relative to both picture-only and text-only formats. Also as predicted, based on the importance of spatial information for object assembly, performance on pictures alone exceeded that based on text alone. This finding contrasts with some previous work demonstrating a slight advantage of text relative to pictures (Gyselinck et al., 2002). However, the difference between Gyselinck et al. (2002) and our findings may be explained as a function of informational consistency between text and pictures. That is, while Gyselinck and colleagues' multimedia presentations did not contain informational consistency across formats, the multimedia stimuli used in the present experiments were developed with a greater degree of informational equivalence.

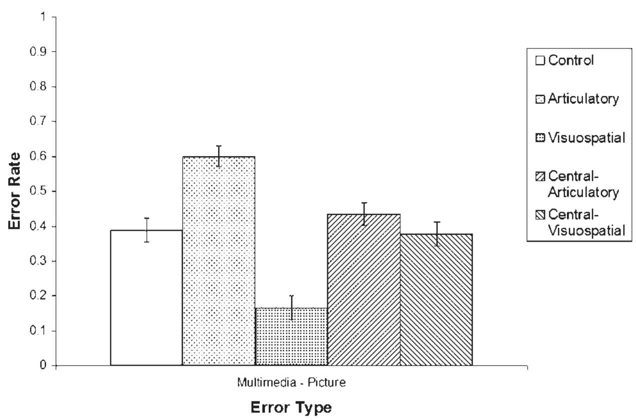

Figure 4. Experiment 1 mean occurrence and standard error for the MM-P error type, for the control, articulatory, visuospatial, and two central secondary task groups

Assembly sequence pictures convey complex spatial relationships inherent to assembly materials (Stone & Glock, 1981), and as such should be more useful for learning when closer in informational content to the corresponding text. This position is further supported by the order verification data. Participants made faster and more accurate order verification judgments as a result of multimedia learning relative to both picture- and text-only formats, and greater accuracy following picture relative to text learning. This task uniquely examines the extent to which mental representations facilitate the retrieval of order information inherent in assembly sequences. It suggests that stronger procedural representations are constructed following multimedia presentations as compared to single-format learning conditions.

Despite the advantage of multimedia learning, the format recall data suggest that participants are not aware of using the text to their benefit. Based on these data, people seem to mainly remember the pictures from the multimedia presentations. This supports the idea that pictures may be the primary information source when object identities are easily derived and spatial relationships are essential to comprehension. Yet multimedia still resulted in better memory for the procedures than picture-only presentations, which suggests participants acquired something additional when text was included.

The current research also examined the role of working memory processes during procedural learning tasks, extending previous subsystem work to also include central executive processes. We used four secondary tasks, two examining the spatial and verbal components of working memory, and two focusing on central executive resources. Although the central-articulatory task clearly has an articulatory component, results of both free recall and order verification accuracy revealed that the articulatory task interfered mainly with text processing and the central-articulatory task interfered mainly with multimedia processing. An analogous pattern was revealed with the visuospatial (interfering with picture processing) and central-visuospatial (interfering with multimedia processing) tasks. The findings for the subsystem tasks are in line with Mayer's (1997) multimedia theory, Baddeley's (1992, 2002) working memory model, and previous findings reported by Gyselinck et al. (2002). According to this work, articulatory suppression should selectively interfere with phonological resources, decreasing the subsequent recall and use of memories acquired during text-only, but not picture-only, and perhaps not multimedia experiences. In contrast, visuospatial suppression should selectively interfere with spatial resources, impairing memory for information presented in picture-only or multimedia format. The central task findings, although not previously studied with multimedia presentations, also fit nicely with Mayer's (1997) model. Central-articulatory and central-visuospatial suppression exclusively interfered with multimedia processing, supporting Mayer's position that central resources are necessary for the integration of texts and pictures.

Additional support for the selective interference of both articulatory and visuospatial resources comes from the format recall results, which revealed higher MM-P error rates following articulatory, and lower MM-P rates following visuospatial, interference. The articulatory task appears to have enhanced the importance of pictures, to the extent that participants were even more likely to misremember multimedia presentations as picture-only. Interestingly, articulatory suppression during multimedia learning did not lead to significant reductions in free recall or order verification, suggesting participants were able to construct memory from the pictures with minimal verbal support. The visuospatial task, in contrast, appears to have increased the importance of text, effectively decreasing the MM-P error rate relative to control group performance. Interestingly, there was no corresponding increase in MM-T error rates with the visuospatial task, suggesting that participants reduced reliance on pictures without increasing their phenomenological memory for the text.

The novel use of a source monitoring task during multimedia learning revealed exceptionally high MM-P misattribution rates across secondary groups. There are at least two explanations for these MM-P effects. First, those who have high misattribution rates are not taking advantage of text during multimedia presentations, which should manifest itself in similar performance for picture and multimedia on free recall and order verification tasks. Second, high MM-P misattribution rates may be a result of participants using the text, but this format has little impact on final memory representations. These data support second explanation: While participants had used the text to their advantage during learning with multimedia, the use had little phenomenological impact. Specifically, an exploratory analysis of learning trials followed by MM-P misattribution errors reveals higher performance on both free recall and order verification tasks following multimedia, relative to picture-only learning. If participants were not using the text component during multimedia presentations, performance on these tasks should be equivalent to that demonstrated following picture-only presentations. We note that MM-P misattributions were found for 70 of the 75 participants in this experiment. Of the 5 who did not misattribute, 4 were in the visuospatial secondary group, suggesting increases in the phenomenological impact of text as a function of decreased reliance on pictures, but not to the extent that MM-T misattributions increased in a significant way.

Overall, the results supported our hypotheses; an advantage for multimedia presentations over text- and picture-only conditions, a general advantage for pictures over texts, selective interference as a function of articulatory, visuospatial, and central suppression, and participants' focus on picture-based material identified through source monitoring errors.

EXPERIMENT 2

Results of Experiment 1 show selective interference with single-format learning and central executive interference with multimedia learning. Experiment 2 examined a different type of interference task, one which requires sequential updating of working memory. This type of interference task is particularly interesting in the context of procedural learning, since this type of learning requires the acquisition of sequential order information. As such, we hypothesize that a secondary sequence task should interfere with procedural sequence learning, regardless of format. It is important to realize, however, that the present secondary task may also require central executive resources such as the updating or monitoring of working memory representations, and temporal tagging (e.g., Miyake & Shah, 1999; Miyake, Friedman, Emerson, Witzki, & Howerter, 2000). This possibility will be further explored in the discussion section.

We incorporated two secondary sequence tasks, one phonologically oriented and one visuospatially oriented, so as to examine selective interference with single-format learning and multimedia learning. While the predictions for these two secondary tasks relative to free recall and order verification performance are identical, they differ with respect to format recall performance. The phonologically-oriented task should increase participants' reliance on pictures during multimedia presentations, consequently increasing MM-P error rates relative to controls. The spatially-oriented task, in contrast, is expected to decrease MM-P error rates due to an increased reliance on texts during multimedia presentations.

Method

Participants

Seventy-five Tufts University undergraduates participated for partial course credit.

Materials

All materials were identical to those used in Experiment 1, with the exception of the secondary tasks.

Secondary tasks. Both secondary tasks, verbal and spatial, required sequence processing and updating. The phonological task was adapted from Rabinowitz, Craik, and Ackerman (1982). Participants listened to a sequence of digits via headphones, monitoring for the occurrence of a target sequence of three consecutive odd digits. The full recording contained 80 target sequences within a 650-digit list, presented at a rate of 1.5 seconds/digit. We constructed the list based on the following restrictions: at least one and no more than five digits could occur between target sequences and no more than two even numbers could occur in a row. The entire list was recorded onto a cassette by a male speaker. The spatial secondary task required participants to listen to a sequence of beeps, presented in either the left or right ear, at a rate of 1.5 seconds/beep, and a duration of 0.5 second/beep. The target sequence, in this case, was three beeps in the left ear. The list construction followed the same constraints as the verbal task (Rabinowitz et al., 1982); at least one and no more than five beeps could occur between target sequences, and no more than two right ear beeps could occur in a row.

Procedure

All procedures were identical to Experiment 1, with the exception of the secondary tasks.

Secondary tasks. Participants were randomly assigned to one of three secondary task groups (control, verbal sequence, or spatial sequence). The control group had no secondary task. Both secondary tasks required participants to monitor a tape recording for a target sequence while simultaneously studying the assembly sequences. To ensure participants paid attention to the secondary task, they were required to make tick marks on a sheet of paper for each identified target sequence. Participants with secondary tasks practiced their task in a session with five target sequences within a 27-digit (or beep) sequence. Participants had to reach criterion during practice, defined as correct identification of all five targets, before beginning the primary task. The two secondary tasks were experimenter-paced, and monitored by participants via headphones.

Results

Analyses

Data from two participants were discarded due to chance (50%) performance on the order verification task. To examine between-groups differences on free recall and order verification performance, three mixed (learning condition: picture, text, multimedia; secondary task: phonological-sequential, visuospatial-sequential) factorial repeated measures ANOVAs were conducted on accuracy and RT data. To investigate between-groups differences of MM-P format recall errors, a repeated measures ANOVA was conducted on data from the two secondary task groups relative to the control group, and planned comparisons further probed for between-group differences.

Learning format advantages

Control group means can be found in Table 1, for free recall and order verification tasks.

Free recall. Analysis revealed an effect of learning format, $F(2, 52) = 30.290$, $p < 0.01$, $MSE = 0.032$. Planned comparisons revealed better recall following multimedia presentations relative to both picture-only ($p < 0.01$, Mean Difference = 0.198) and text-only ($p < 0.01$, Mean Difference = 0.383) formats. Recall rates following the picture-only format exceeded those following the text-only format ($p < 0.01$, Mean Difference = 0.185).

Order verification. Analysis of order verification accuracy revealed a learning format effect, $F(2, 52) = 25.709$, $p < 0.01$, $MSE = 0.009$. Planned comparisons revealed higher accuracy following multimedia relative to both picture-only ($p < 0.01$, Mean Difference = 0.08) and text-only ($p < 0.01$, Mean Difference = 0.185) formats. Additionally, participants had greater accuracy following picture-only relative to text-only presentations ($p < 0.01$, Mean Difference = 0.104).

Analysis of order verification RT also revealed a learning format effect, $F(2, 52) = 27.670$, $p < 0.01$, $MSE = 0.802$. Planned comparisons revealed faster responses following multimedia relative to both picture-only ($p < 0.01$, Mean Difference = −1.634) and text-only ($p < 0.01$, Mean Difference = −1.853) formats. An additional comparison revealed no difference between picture-only and text-only formats ($p > 0.05$, Mean Difference = −0.219).

Secondary sequential tasks and procedural learning

Free recall and order verification. The secondary tasks resulted in broad-based rather than selective interference, as seen in the significant main effect of secondary task for both recall, $F(2, 72) = 9.068$, $p < 0.01$, $MSE = 0.0958$, and order verification accuracy,

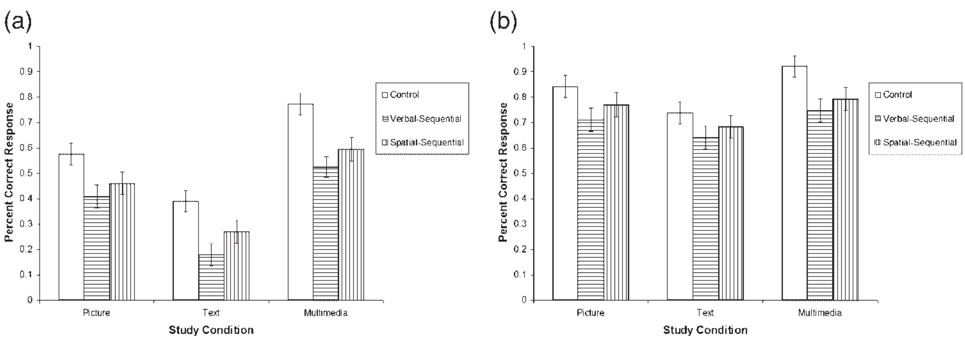

Figure 5. a–b. Experiment 2 mean control, phonological- and visuospatial-sequential group accuracy rates and standard error derived from the free recall (a) and order verification (b) tasks for picture, text, and multimedia conditions

$F(2, 72) = 10.473$, $p < 0.01$, $MSE = 0.0336$. The control group had higher overall accuracy compared to both secondary task groups (see Figures 5a and 5b for recall and order verification, respectively). Neither task showed a learning format by secondary task interaction, as would be expected for selective interference.

Source monitoring and multimedia. An omnibus analysis of format recall error rates revealed a main effect of error type, $F(5, 360) = 101.81$, $p < 0.01$, $MSE = 0.027$, replicating Experiment 1 results. Planned comparisons revealed higher MM-P error rates relative to all other types (all $ps < 0.01$); T-P ($M = 0.058$, $SD = 0.124$), T-MM ($M = 0.062$, $SD = 0.128$), P-T ($M = 0.038$, $SD = 0.08$), P-MM ($M = 0.093$, $SD = 0.14$), MM-T ($M = 0.044$, $SD = 0.107$), and MM-P ($M = 0.524$, $SD = 0.343$). A one-way ANOVA of MM-P format recall errors revealed a significant main effect of secondary task, $F(2, 74) = 3.905$, $p < 0.05$, $MSw = 0.109$. Planned comparisons revealed lower MM-P error rates in the control group relative to both secondary task groups, $t(72) = 2.781$, $p < 0.01$ (see Figure 6).

Discussion

Control group performance replicated the multimedia effects for both recall and order verification seen in Experiment 1. Specifically, performance on these tasks was better following multimedia relative to both picture- and text-only presentations.

As predicted, secondary tasks requiring sequence updating in working memory produced broad-based performance decrements on free recall and order verification performance, suggesting that the secondary task tapped into a common processing system. This secondary task was specifically chosen because it requires sequence monitoring and updating, two working memory processes hypothesized to operate during procedural sequence learning. The results largely support this hypothesis, by dissociating central executive sequential monitoring, updating, and temporal tagging processes from both the repetitive (syllable string repetition, finger tapping; articulatory and visuospatial, respectively) and random (central executive) processes used in Experiment 1. This result suggests the ability to interfere selectively with two central executive functions: switching between and integrating dual-format information as seen in Experiment 1 (Monsell, 1996), and monitoring, updating, and temporal tagging in sequential tasks as seen in Experiment 2

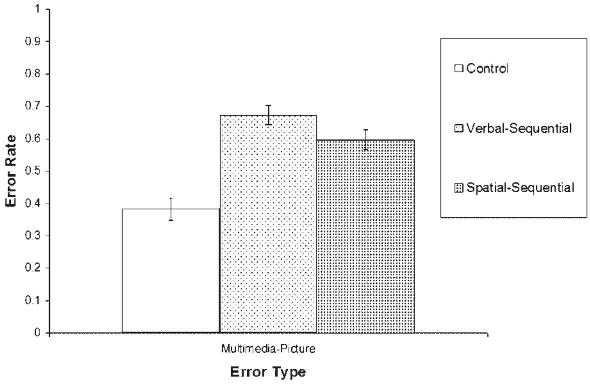

Figure 6. Experiment 2 mean occurrence and standard error for the MM-P (misattributing multimedia learning as picture-only) format recall test error type for control, phonological- and visuospatial-sequential groups

(Miyake & Shah, 1999; Miyake et al., 2000). Interestingly, the phonological or spatial orientation of the sequential secondary tasks did not differentially interfere with specific learning formats, suggesting that sequential encoding may occur independently of format encoding. Congruent with this, and counter to our initial hypothesis, MM-P error rates increased for both phonological and spatial sequence tasks, suggesting no specific processing subsystem orientation for these secondary tasks. These results support the notion that complex procedural sequence processing within working memory requires the monitoring and updating of sequentially presented steps, recruiting similar working memory resources as more basic sequence processing tasks.

It must be noted that Experiments 1 and 2 used self- and experimenter-paced secondary tasks, respectively. The non-selective interference seen in Experiment 2 could have been partially a result of secondary task difficulty, with experimenter-paced monitoring being more difficult (and perhaps less automatized) than the self-paced production in Experiment 1. In addition, the Experiment 2 secondary tasks required a response to target sequences, an additional component that may have further taxed working memory relative to the tasks employed in Experiment 1. It would be useful to employ a broader range of secondary tasks that better control difficulty level, allowing for a more conclusive evaluation of working memory processes during procedural multimedia learning. Further, comparing secondary tasks within- rather than between-subjects may better assess both the selectivity and the relative difficulty of these tasks, and better consider such variables as individual spatial and verbal ability (e.g., Gyselinck et al., 2002). Thus, while the present results are promising in that they implicate selective and global working memory functions during the learning of procedures, they also highlight several avenues for future research.

GENERAL DISCUSSION

Experiments 1 and 2 clearly identified multimedia as a useful presentation format for learning procedural sequence information. This result extends Mayer's (1997) previous work to materials placing an emphasis on sequential details. In addition, coupling pictures with text appears to be more effective than presenting pictures alone. This effect may be due at least partially to the importance of text in conveying object identities, particularly in the case of novel, complex, important, or strange items (such as an axle or a ship's mast). If indeed text plays this role in procedural learning from multimedia, it would be appropriate to assess the role of text-based labels on picture-based learning sequences. Such labels would provide additional identity information (e.g., Graesser & Olde, 2003), critical to the comprehension (and of course, identification) of such presentations.

These results also speak to the benefits of pictures for spatially-rich presentations and aligns with the view that pictures are especially useful for depicting spatial relationships described in a text (Levie & Lentz, 1982; Stone & Glock, 1981), and contrasts with the text benefits associated with expository presentations (e.g., Gyselinck et al., 2002). Thus, to the extent that pictures emphasize spatial relationships, which certainly occur in particular types of expository texts (e.g., scientific and technical descriptions of a bicycle pump; Mayer & Sims, 1994), or that are inherent to most types of procedural sequences (e.g., putting together furniture or a building block toy), pictures may be especially useful.

The present results also provide strong support for Baddeley's (1992, 2003) working memory model, and relatedly, Mayer's (1997) conceptualization of dual-coding theory as it applies to multimedia presentations. The present experiments demonstrate the dissociable nature of subsystem (i.e., visuospatial sketchpad, phonological loop) and central (i.e., central executive) working memory resources. In these experiments, a secondary task recruiting articulatory processes selectively interfered with text processing, while a secondary task recruiting visuospatial processes selectively interfered with picture processing, including picture-focused processing with multimedia. The selective nature of this interference provides support for the role of these systems in procedural learning. Additionally, the investigation of secondary tasks recruiting central executive resources, regardless of whether they also involve articulatory or visuospatial subsystems, demonstrated selective interference with multimedia processing. Specifically, random generation tasks interfered with attention switching and the integration of pictures and text, and sequential monitoring and updating tasks interfered with assembly sequence processing at a more global level. These results speak not only to the validity of Baddeley's model, but also to the processing of procedural single- and dual-format presentations, which appears to be done in a manner similar to that of materials not requiring sequential encoding. We also consider our results novel as they uniquely demonstrate interference with sequential encoding, suggesting independence of format-specific and sequence processing working memory mechanisms.

The present experiments provide evidence that while multimedia presentations may impart benefits for memory, users may be relatively unaware of those benefits. The source monitoring results showed that participants often incorrectly identified multimedia presentations as picture-based only. At some level this suggests that individuals may intuitively feel that pictures are more important or effective than text for understanding construction tasks like those provided in these experiments. This does not mean that participants actively ignore text, but rather that in retrospect, they find the pictures more readily identifiable as useful for completing procedural tasks. One implication of such findings is that instructions may need to readily focus users' attention to text when it is critically important, or when it conveys information that the pictures could not include. For example, textbooks that include pictures may necessitate more than just a simple caption to ensure that readers fully understand figures, graphs, and images. Indeed, later recall of textbook information may be influenced by expectations of what a student read. A second implication is that evaluations of the effectiveness of multimedia may be biased towards picture-specific comments, potentially undervaluing the benefits of included text. Thus, assessments of educational presentations or commercial advertisements that rely on think-aloud procedures or value judgments may underestimate the contributions of features (i.e., text) that consumers seemingly appear to ignore. Indeed, the source monitoring effects we report here suggest that pictures may be seductive in the degree to which they lead individuals to rely on them, however appropriate, for information delivery (e.g., Garner, Gillingham, & White, 1989; Harp & Mayer, 1998). This is not a problem if the pictures and text convey roughly the same information; when pictures and texts convey complementary information, however, comprehension may suffer.

In fact, in order to make valid comparisons across experimental conditions involving text-only, picture-only, and multimedia presentations, as we have attempted to do here, there must be informational equivalence, to the extent possible, across conditions. While every effort was made to derive text information directly and comprehensively from the pictures, even in this context pictures may inherently provide more detailed information or slightly different information than text, especially with regard to object colors, shapes, contours, and depth. However, it must be noted that order verification performance (accuracy and RT) was derived from both picture-only and text-only testing trials, for all learning conditions. Surprisingly, pilot research (Brunyé, Taylor, & Rapp, 2003) has shown fairly equivalent performance on both trial types, suggesting that text learning can be easily applied to picture testing, and vice-versa. If the text did not provide a solid foundation for effective visualization, the results should reflect this distinction by demonstrating relatively poor picture trial (vs. text trial) performance following text-only learning. However, they do not.

An additional concern found in most multimedia experiments involves comparison across experimental conditions. To the extent that multimedia is more effective than picture- or text-only presentations, this effect can be attributed either to the repetition inherent to multimedia *or* an actual format effect. The present experiments attempted to control for repetition by presenting each single format step twice simultaneously. We cannot be certain that participants took advantage of this repetition, therefore leaving open the possibility for repetition effects in the multimedia case, perhaps complementing the format effects. However, we also cannot be certain that people both viewed the picture and read the text in the multimedia case. Source monitoring results suggest minimal text viewing, yet memory performance suggests a multimedia advantage. Future work should assess these issues using eye-tracking methodologies to evaluate exactly where participants look as they learn procedural sequences.

Despite these concerns, the present research highlights several important aspects of multimedia learning, particularly with regard to procedural sequences. While the majority of these results (multimedia advantages, selective interference) fit with working memory and multimedia theory, some results (source monitoring errors) demand additional research. In addition, while sequential memory is an important research area, it would be interesting to apply the present paradigm to actual object assembly tasks. Object assembly as a dependent measure could greatly inform the present results, extending their applicability to real-world tasks.

Procedural presentations, whether observed on a television cooking program, outlined in a planning pamphlet from a home improvement store, detailed in a software owner's guide, or studied from the instructions for a construction toy, often succeed or fail for a variety of reasons. Sometimes these presentations are effective and understandable, and we can enjoy the fruits of our labors (as, for example, not only the chocolate from a Kinder Egg, but also a working version of the toy inside); other times, though, these presentations are less than effective. One critical reason for failures in the design of such presentations involves author ignorance or avoidance of cognitive principles of learning (Zacks & Tversky, 2003). For example, pictures may present inaccurate depictions of important concepts, or text may understate or ambiguously state the steps necessary. Current work on the impact of multimedia seeks to apply cognitive theories towards an understanding of the presentation conditions that foster and impede encoding and retrieval in memory. Thus, we believe these findings have potential relevance not just for existing theories of learning, but also for real-world application in the design of effective procedural instructions.

ACKNOWLEDGEMENTS

We thank Alexander Field, Kristin Saldarelli, Shana Horak, Julia Green, Alan Manos and Heather Kenney for their diligent assistance in data collection. Special thanks to Drs George Wolford and Michael Carlin for their advice regarding statistical analyses. We are also grateful to Drs Francesca Pazzaglia, Caroline Cao, and several anonymous reviewers for their insightful comments on earlier versions of the article.

REFERENCES

Allen, W. H. (1967). Media stimulus and types of learning. *Audiovisual Instruction, 12*, 27–31.

Allen, W. H. (1971). Instructional media research: Past, present, and future. *AV Communications Review, 19*, 5–18.

Baddeley, A. D. (1992). Working memory. *Science, 255*, 556–559.

Baddeley, A. D. (1996). Exploring the central executive. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 49*, 5–28.

Baddeley, A. D. (2002). Is working memory still working? *European Psychologist, 7*, 85–97.

Baddeley, A. D., Emslie, H., Kolodny, J., & Duncan, J. (1998). Random generation and the executive control of working memory. *Quarterly Journal of Experimental Psychology, 51A*, 818–852.

Baggett, P. (1979). Structurally equivalent stories in movie and text and the effect of the medium on recall. *Journal of Verbal Learning & Verbal Behavior, 18*, 333–356.

Brunyé, T., Taylor, H. A., & Rapp, D. N. (2003). When do you put on Snoopy's hat? Influences of presentation modality on memory for procedural instructions. Paper Presentation at the 44th annual meeting of the Psychonomics Society, Vancouver, Canada.

De Beni, R., Pazzaglia, F., Gyselinck, V., & Meneghetti, C. (2005). Visuospatial working memory and mental representation of spatial descriptions. *European Journal of Cognitive Psychology, 17*, 77–95.

Diehl, V. V., & Mills, C. B. (1995). The effects of interaction with the device described by procedural text on recall, true/false, and task performance. *Memory and Cognition, 23*, 675–688.

Farmer, E. W., Berman, J. V., Fletcher, Y. L. (1986). Evidence for a visuo-spatial scratchpad in working memory. *Quarterly Journal of Experimental Psychology, 38*, 675–688.

Garner, R., Gillingham, M. G., & White, C. S. (1989). Effects of 'seductive details' on macroprocessing and microprocessing in adults and children. *Cognition and Instruction, 6*, 41–57.

Garrison, W. T. (1978). The context bound effects of picture-text amalgams: Two studies. Dissertation Abstracts International, *39*, 4137A.

Glenberg, A. M., & Langston, W. E. (1992). Comprehension of illustrated text: Pictures help to build mental models. *Journal of Memory and Language, 31*, 129–151.

Graesser, A. C., & Olde, B. A. (2003). How does one know whether a person understands a device?: The quality of the questions the person asks when the device breaks down. *Journal of Educational Psychology, 95*, 524–536.

Gyselinck, V., Cornoldi, C., Dubois, V., De Beni, R., & Ehrlich, M. F. (2002). Visuospatial memory and phonological loop in learning from multimedia. *Applied Cognitive Psychology, 16*, 665–685.

Gyselinck, V., & Tardieu, H. (1999). The role of illustrations in text comprehension: What, when, for whom, and why? In H. van Oostendorp,& S. R. Goldman (Eds.), *The construction of mental representations during reading* (pp. 195–218). Mahwah, NJ, US: Lawrence Erlbaum Associates.

Harp, S. F., & Mayer, R. E. (1998). How seductive details do their damage: A theory of cognitive interest in science learning. *Journal of Educational Psychology, 90*, 414–434.

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin, 114*, 3–28.

Kozma, R. B. (1991). Learning with media. *Review of Educational Research, 61*, 179–212.

Kruley, P., Sciama, S. C., & Glenberg, A. M. (1994). On-line processing of textual illustrations in the visuospatial sketchpad: Evidence from dual-task studies. *Memory & Cognition, 22*, 261–272.

Levie, W. H., & Lentz, R. (1982). Effects of text illustrations: A review of research. *Educational Communication & Technology Journal, 30*, 195–232.

Logan, G. D. (1985). Executive control of thought and action. *Acta Psychologica, 60*, 193–210.

Logie, R. H. (1995). *Visuo-spatial working memory*. Hove, UK: Lawrence Erlbaum Associates, Ltd.

Marcus, N., Cooper, M., & Sweller, J. (1996). Understanding instructions. *Journal of Educational Psychology, 88*, 49–63.

Mayer, R. E. (1989). Systematic thinking fostered by illustrations in scientific text. *Journal of Educational Psychology, 81*, 240–246.

Mayer, R. E. (1997). Multimedia learning: Are we asking the right questions? *Educational Psychologist, 32*, 1–19.

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

Mayer, R. E., & Anderson, R. B. (1991). Animations need narrations: An experimental test of a dual-coding hypothesis. *Journal of Educational Psychology, 83*, 484–490.
Mayer, R. E., Bove, W., Bryman, A., Mars, R., & Tapangco, L. (1995). A generative theory of textbook design: Using annotated illustrations to foster meaningful learning of scientific text. *Educational Technology Research and Development, 43*, 31–44.
Mayer, R. E., & Gallini, J. K. (1990). When is an illustration worth ten thousand words? *Journal of Educational Psychology, 82*, 715–726.
Mayer, R. E. & Moreno, R. (2002). Aids to computer-based multimedia learning. *Learning and Instruction, 12*, 107–119.
Mayer, R. E., & Sims, V. K. (1994). For whom is a picture worth a thousand words? Extensions of a Dual-coding theory of multimedia learning. *Journal of Educational Psychology, 86*, 389–401.
Miyake, A., & Shah, P. (1999). Toward unified theories of working memory: Emerging general consensus, unresolved theoretical issues, and future research directions. In A. Miyake, & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 442–481). New York: Cambridge Univ. Press.
Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., & Howerter, A. (2000). The unity and diversity of executive functions and their contribution to complex 'frontal lobe' tasks: A latent variable analysis. *Cognitive Psychology, 41*, 49–100.
Monsell, S. (1996). Control of mental processes. In V. Bruce (Ed.), *Unsolved mysteries of the mind: Tutorial essays in cognition* (pp. 93–148). Hove, UK: Erlbaum.
Novick, L. R., & Morse, D. L. (2000). Folding a fish, making a mushroom: The role of diagrams in executing assembly procedures. *Memory & Cognition, 28*, 1242–1256.
Paivio, A. (1965). Abstractness, imagery, and meaningfulness in paired-associate learning. *Journal of Verbal Learning & Verbal Behavior, 4*, 32–38.
Paivio, A. (1986). *Mental representations: A dual coding approach*. Oxford, England: Oxford University Press.
Peeck, J. (1994). Enhancing graphic-effects in instructional texts: Influencing learning activities. In W. Schnotz, & R. W. Kulhavy (Eds.), *Comprehension of graphics: Advances in psychology* (pp. 291–301). Amsterdam, Netherlands: North-Holland/Elsevier Science.
Peeck, J., & Jans, M. W. (1987). Delayed retention of orally presented text with pictorial support. *British Journal of Educational Psychology, 57*, 412–416.
Rabinowitz, J. C., Craik, F.I., & Ackerman, B. P. (1982). A processing resource account of age differences in recall. *Canadian Journal of Psychology, 36*, 325–344.
Slamecka, N. J., & Graf, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning and Memory, 6*, 592–604.
Smith, E. E., & Jonides, J. (1999). Storage and executive processes in the frontal lobes. *Science, 283*, 1657–1661.
Soraci, S. A., Franks, J. J., Bransford, J. D., Chechile, R. A., Belli, R. F., Carr, M., & Carlin, M. (1994) Incongruous item generation effects: A multiple-cue perspective. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 20*, 67–78.
Stone, D. E., & Glock, M. D. (1981). How do young adults read directions with and without pictures? *Journal of Educational Psychology, 73*, 419–426.
Wittrock, M. C. (1989). Generative processes of education. *Educational Psychologist, 24*, 345–376.
Zacks, J. M, & Tversky, B. (2003). Structuring information interfaces for procedural learning. *Journal of Experimental Psychology: Applied, 9*, 88–100.
Zwaan, R. A. (1994). Effect of genre expectations on text comprehension. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 20*, 920–933.

