Applied Cognitive Psychology, Appl. Cognit. Psychol. **30**: 1052–1060 (2016)
Published online in Wiley Online Library (wileyonlinelibrary.com) **DOI**: 10.1002/acp.3299

Procedural Multimedia Presentations: The Effects of Working Memory and Task Complexity on Instruction Time and Assembly Accuracy

NATALIA IRRAZABAL[1]*, GASTÓN SAUX[2] and DEBORA BURIN[3]

[1]*National Council of Scientific and Technical Research (CONICET), Palermo University, Buenos Aires, Argentina*
[2]*National Council of Scientific Technical Research (CONICET)Buenos Aires, Argentina*
[3]*National Council of Scientific and Technical Research (CONICET), University of Buenos Aires, Buenos Aires, Argentina*

Summary: Procedural text conveys information of a series of steps to be performed. This study examined the role of verbal and visuo-spatial WM in comprehension and execution of assembly instructions, as a function of format (text, images, multimedia) and task complexity (three or five steps). One hundred and eight participants read and executed 27 instructions to assemble a LEGO™ object, in single and dual task conditions. Study times and errors during assembly were measured. Participants processed faster pictorial and multimedia instructions than text instructions, and made fewer errors in the execution of multimedia instructions. Dual task affected more text or picture-only, than multimedia presentation. A verbal secondary task caused more errors in text or picture-only presentations, and spatial secondary task also caused interference in text-only instructions. Overall, these results support the multimedia advantage, and the role of both verbal and visuo-spatial WM, when understanding instructions.

Understanding a sequence of instructions implies comprehending logical and pragmatic relations between the procedural steps in order to achieve a certain result (Brunyé, Taylor, Rapp, & Spiro, 2006). This is done by constructing a coherent and integrated mental model, in which goals, sub-goals, and actions are strategically represented and updated in an execution-oriented representation (Diehl & Bergfeld Mills, 2002; Geiger & Millis, 2004; Mills, Diehl, Birkmire, & Mou, 1995). Comprehending and executing procedures is important for various everyday activities, such as fulfilling tasks at work, preparing a new recipe, or following a medical prescription. Procedural instructions also include the specific case of assembling objects (e.g. LEGO*™* objects, kitchen appliances, home furniture, puzzles, building houses, among others). Therefore, several features such as how and when the information is presented, the complexity of the relations between states and goals, and the degree of implicit ideas involved in the instruction may impact on its comprehension and execution.

MULTIMEDIA EFFECT IN PROCEDURAL TEXT

In everyday life, most instructions are illustrated by graphics. In experimental and educational settings, comprehension and learning from text benefit from images (e.g. Mayer, 2001, 2014), an overall effect called the 'multimedia effect'. Multimedia includes any presentation that combines more than one format (text and image), in a single sensory modality (auditory or visual) or in a combination of them (Mayer, 2001). In expository texts, the multimedia effect and other aspects of multimedia presentations have been extensively analyzed (Mayer, 2014).

Studies addressing the multimedia effect in procedural texts have also found an overall benefit for the multimedia conditions (text and pictures), in comparison with text-only conditions (Brunyé et al., 2006; Marcus, Cooper, & Sweller, 1996; Michas & Berry, 2000; Mills et al., 1995; Novick & Morse, 2000; van Genuchten, van Hooijdonk, Schüler, & Scheiter, 2014; Zacks & Tversky, 2003). For example, Brunyé et al. (2006) examined the association between format of instruction presentation and memory, and found better recall of instructions in the multimedia condition. However, this line of research has seldom examined the effects on actual assembly, as opposed to memory (retention vs. application of knowledge, Michas & Berry, 2000; knowing 'how' vs. knowing 'that', van Genuchten et al., 2014), which would be relevant considering real object assembly tasks in everyday life. Novick and Morse (2000) showed that a diagram depicting the final result of the instruction led to better accuracy in execution in short verbal instructions (five steps); whereas in longer instructions (16 steps), step-by-step diagrams were more effective than a final state graph. Michas and Berry (2000) had participants learn a procedural task (first-aid bandage), through text, line drawings, video, or video stills. Overall, they found an advantage for combined text and picture, and animated presentations, over single format ones, for both memory and performance. For their part, Brunyé, Taylor, and Rapp (2007) analyzed both recall (free recall and order verification task) and assembly of Kinder Egg™ toys as a function of format of instruction presentation (verbal, pictorial, and multimedia). For assembly, similar results were obtained with multimedia and picture-only instructions. In sum, research about the effects of format of instructions on memory and assembly performance is scarce and has found an advantage of multimedia presentations (Brunyé et al., 2006, 2007; Michas & Berry, 2000; Novick & Morse, 2000; van Genuchten et al., 2014; Zacks & Tversky, 2003), or similar results with multimedia and picture-only presentations (Brunyé et al., 2007).

*Correspondence to: Natalia Irrazabal, National Council of Scientific and Technical Research (CONICET) - Palermo University, Mario Bravo 1259, (1175ABW) Buenos Aires, Argentina.
E-mail: nirrazabal@psi.uba.ar

MULTIMEDIA AND WORKING MEMORY IN INSTRUCTIONS

Construction and maintenance of a mental model of instructional material rely on working memory (WM) (Mayer, 2001, 2014). This system is engaged in temporary maintenance and simultaneous processing of information required in complex cognitive activities (Baddeley, 2002). In Baddeley's classic model (Baddeley, 1986, 2002), WM is conceived as a multi-component temporary storage and processing system: two modality specific storage and maintenance subsystems (verbal and visuo-spatial), limited in capacity and duration, and a central executive component responsible for coordination, attentional control, and retrieval from long term memory. A fourth component added later (Baddeley, 2000, 2002) would deal with higher-order, multidimensional representations, linking the more specialized maintenance subsystems with long-term episodic memory. Given that all subsystems are assumed to be limited in the amount of information they can handle in a given moment (Baddeley, 2002), the dual task logic is applied to explore WM components: a secondary task is performed along a primary one, manipulating presentation modality, amount of items, or complexity of the tasks, and registering changes in processing time or accuracy as a result. A typical result is that if both tasks do not compete for a particular component, and the number of to-be-remembered items does not exceed capacity, more time is needed but accuracy does not suffer. In particular, dual task paradigms have been employed to assess whether a primary task relies on verbal or visuo-spatial WM components (e.g. Baddeley, 1986, 2002). In the WM literature, simple secondary tasks such as articulatory suppression (Baddeley, 1986; Murray, 1968) and spatial tapping (Baddeley, 1986; Baddeley & Lieberman, 1980) are assumed to interfere with WM modality specific subsystems, whereas more complex secondary tasks (e.g. random generation, Baddeley, 2002) are supposed to rely upon central executive resources. For example, Gyselinck, Jamet, and Dubois (2008) reported evidence that a verbal concurrent task impaired the comprehension of complex documents (thus suggesting that the verbal WM plays an important role in text comprehension), whereas spatial concurrent tasks affected performance as soon as visuo-spatial processing was required (e.g. in multimedia presentations or in tasks that demand mental imagery). Also, when secondary tasks become more complex, central executive involvement cannot be ruled out (Baddeley, 1986, 2002).

From the initial dual coding hypothesis (Paivio, 1971, 1986), to more recent multimedia processing models such as the Cognitive Theory of Multimedia Learning (Mayer, 2001, 2014) and the Cognitive Load Theory (Sweller, van Merriënboer, & Paas, 1998; Mayer & Moreno, 2003), the multimedia effect has been attributed to different representational systems for verbal and pictorial information. Instructional materials are processed and integrated in the limited WM space; this posits a limited load that the learner can process in a given moment (Sweller et al., 1998; Mayer & Moreno, 2003). Cognitive load can be reflected by performance in terms of time-on-task and accuracy (Brunken, Plass, & Leutner, 2003; Sweller et al., 1998). The multimedia benefit would be attributed to activation of visuo-spatial and verbal WM. For example, Gyselinck, Cornoldi, Dubois, De Beni, and Ehrlich (2002) showed that both WM components collaborate in multimedia comprehension of scientific texts. They found that a concurrent visuo-spatial task eliminated the beneficial effect of multimedia condition, and affected the comprehension of pictures presented on their own, whereas a concurrent verbal task impaired performance in text-only and multimedia condition. Employing both verbal and visuo-spatial WM to build and maintain an integrated representation could reduce cognitive load, given certain conditions such as contiguity, coherence, and redundancy (see Mayer, 2001, 2014; Mayer & Moreno, 2003). For example, sequential integration of a series of steps taxes WM more than simultaneous presentations. The multimedia benefit could stem from repetition, depicting the same information in both formats, or because they provide complementary information leading to 'active integration' processing (Brunyé et al., 2007; Mayer, 2001, 2014; Mayer & Moreno, 2003).

In the specific case of assembly instructions, Brunyé et al. (2006) and van Genuchten et al. (2014) addressed the issue of WM involvement in multimedia presentations with selective dual task conditions. Brunyé et al. (2006) examined memory for instructions after text, image, or multimedia presentations, performed alone or combined with articulatory suppression, spatial tapping, or central executive tasks (random generation). Results supported a multimedia advantage over picture-only and text-only presentations, and image advantage over textual instructions. They also revealed selective WM interferences with different procedural formats: articulatory suppression affected text-only whereas spatial tapping interfered with picture-only presentations, and central executive tasks affected multimedia processing. In addition, participants misattributed their learning to the picture-only presentations, thus suggesting image-based internal representations. Meanwhile, van Genuchten et al. (2014) asked participants to read and perform a series of first aid procedural steps, in textual format or multimedia, and with a spatial dual concurrent task (tapping) or without a dual task. They found that the spatial concurrent task interfered with immediate recall and performance accuracy for text-only instructions, but not for multimedia presentations. Given that the spatial secondary task interfered with the text-only presentation, as different from previous research that obtained only intra-modality interference, they argued that their task required generating visual images from text suggesting that visuo-spatial WM would be strongly involved in procedural comprehension. This study would also suggest that text-only instructions further require image generation or construction, increasing cognitive load.

THE PRESENT STUDY

To sum up, previous research has shown that, for both recall of instructions and procedural performance, multimedia or text-only presentations are more efficient than verbal ones (Brunyé et al., 2006, 2007; van Genuchten et al., 2014).

However, there is disagreement regarding WM involvement when following instructions: Brunyé et al. (2006) found format specific dual task interference effects, and central executive involvement for multimedia presentations, whereas van Genuchten et al. (2014) found cross-format interference effects: a spatial concurrent task interfered with text-only presentations. Furthermore, a direct comparison with a verbal secondary task was carried on in only one study (Brunyé et al., 2006). In addition, another cognitive load factor, relative complexity of procedural instructions, as for example the number of sequential steps (Novick & Morse, 2000), has not been systematically varied in previous WM studies. Another limitation is their between-subjects design, with around 20 participants per condition (Brunyé et al., 2006; van Genuchten et al., 2014).

Thus, the present study sought to replicate and extend previous studies regarding presentation format, domain-specific WM components role, and task complexity in procedural instructions performance. We examined comprehension and execution of instructions as a function of presentation format (text, image, and multimedia), concurrent task (verbal, spatial, or no-task), and task complexity (assembly instructions in three or five steps). As different from previous studies, a within subjects design was employed. To analyze the role of verbal and visuo-spatial WM components, we chose secondary tasks previously employed as WM specific interferences, articulatory suppression, and spatial tapping (Baddeley & Lieberman, 1980; Murray, 1968).

Dependent variables included accuracy of execution (assembling errors), and study times during comprehension.

In the first place, we expect better performance in assembly accuracy for multimedia presentations (multimedia effect). In addition, if image-based representations are crucial for understanding instructions (Brunyé et al., 2007; van Genuchten et al., 2014), conditions in which instructions include a pictorial format (picture-only or multimedia) will exhibit fewer errors as compared to text-only presentations. Regarding WM, considering that concurrent tasks and increased number of steps tax WM capacity, we predict main effects of these factors, as well as an interaction between them and the presentation format of the instructions. As in Brunyé et al. (2007) and van Genuchten et al. (2014), we expect that the concurrent tasks (verbal and spatial) will impair performance in single-format presentations (text and picture-only conditions) more than in the multimedia condition.

In addition to performance execution outcomes, the present experiment has also included measures of study time, indexed as mean time per each step of instructions. Study time is a standard technique employed in psycholinguistic research to explore processing during reading comprehension (Aaronson & Ferres, 1984; Just & Carpenter, 1980; Just, Carpenter, & Woolley, 1982); complex WM paradigms with self-paced administrations (as opposed to standard experimenter-controlled procedures) have been employed in the same way. For example, longer viewing times have been interpreted as more rehearsal or elaboration of the to-be-remembered items (Engle, Cantor, & Carullo, 1992; Friedman & Miyake, 2004). Therefore in the present research, study time has been interpreted as processing in WM during learning. According to the literature, less cognitive load would be expected in multimedia (vs. text-only and picture-only formats), three-steps (vs. five-steps), and single task (vs. dual tasks) conditions. This would lead to predict differential study times for these conditions; study times are predicted to be longer as a function of increased load.

In both, accuracy and study times, particular combinations of presentation format and specific WM concurrent tasks, an open question would be whether intra-modal (Brunyé et al., 2006) or cross-modal (van Genuchten et al., 2014) interferences are observed.

METHOD

Participants

One hundred and eight undergraduate psychology students (age $M = 19.88$, $SD = 3.21$) volunteered to participate in exchange for course credits. Participants signed an informed consent form before taking part in the experiment and were debriefed after completing the procedure.

Materials and design

Assembly instructions

Twenty seven instructions consisting of series of steps to assemble a LEGO^{TM} object were employed. For each instruction, the complete object was achieved after three or five steps. Each step instructed on how to put together two parts of the object, until the object was completely assembled. Instructions were presented on a computer screen, using a self-paced method. All sequences were presented using E-prime 2.0, at 300×300 pixel resolution, with 14-point Times New Roman font. Steps were presented one at a time, clearing the screen when the step was accomplished.

Complexity of instructions was defined as the number of steps to achieve the object (three or five steps). The number of steps was implemented after Brunyé et al. (2006, 2007), who used five steps, and Novick and Morse (2000), who employed 5 and 16 steps to fold an origami, but with only two to six items in total, and with the instructions on view.

Instructions were presented in three formats: (i) text-only; (ii) picture-only; and (iii) multimedia (picture + text). A pilot study with 40 college students, in which participants were asked to describe the images, was carried out to validate the instructions' design. In the pilot study, the task was to create text descriptions equivalent to the pictorial instructions. This also ensured that images could be correctly understood. Following Brunyé et al. (2006), we showed each participant the pictorial steps and asked them to generate a written descriptive sentence for each one. Then, three judges read and combined the verbal descriptions to create a concise and coherent text for each step.

In the present experiment, each participant saw 27 instructions in some combination of presentation format, concurrent task, and number of steps. In all conditions, screen was split horizontally into two segments. Picture-only condition presented instructions in pictorial format (an image with a picture of LEGO^{TM} pieces and arrows showing how to assemble the pieces) duplicated on the screen, one in each screen segment (see Figure 1). Text-only condition presented instructions in verbal format duplicated on the screen (one in each screen segment, see Figure 2). Multimedia condition presented instructions combining text and image, each format in each screen segment (see Figure 3). Both screen segments provided the same information twice, simultaneously, in order to control the repetition inherent in multimedia (Brunyé et al., 2006, 2007).

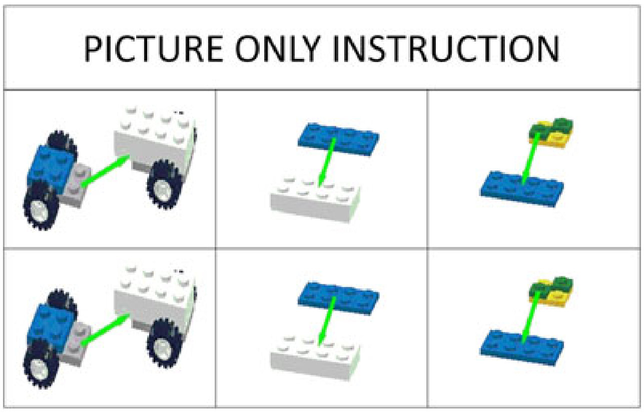

Figure 1. Example of picture-only instructions (three-steps, sequential)

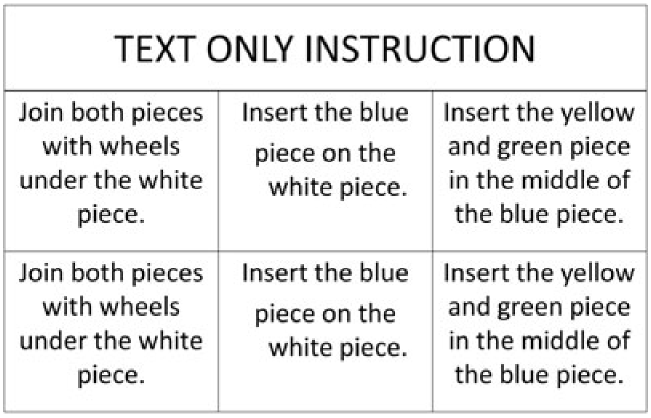

Figure 2. Example of text-only instructions (three-steps, sequential)

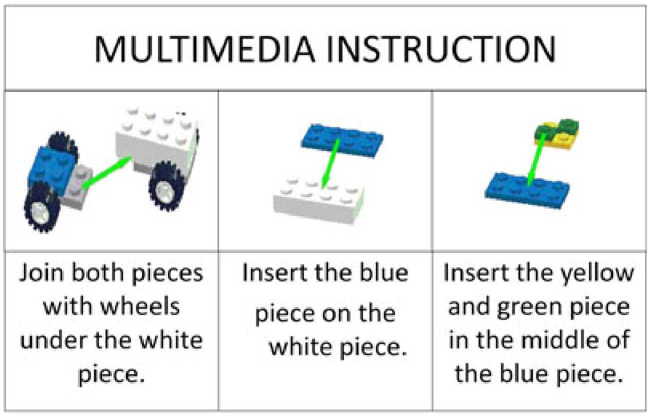

Figure 3. Example of multimedia instructions (three-steps, sequential)

Concurrent tasks

Two secondary tasks were used: articulatory suppression, which consisted of repeating the nonsense syllable 'blah blah' (Murray, 1968); and spatial tapping, in which participants had to tap sequentially, with the left or no dominant hand, the four corners of a wooden surface with wood marks for taps (Baddeley & Lieberman, 1980).

Dependent measures

The present experiment included mean study times during instruction, and mean error rates during assembly. The first variable was calculated as the mean time per screen (i.e. per step) per condition. For the second one, total errors, based on mistakes of sequence or position, were used to calculate the mean error rate corrected by number of steps, per condition.

PROCEDURE

In individual sessions, participants were asked to 'watch and try to remember series of steps, in order to assemble LEGO^{TM} objects'. First, they completed nine training instructions, with supervision and feedback from a research assistant. At the beginning, they practiced instructions in each format without a concurrent task. Then, they trained with each concurrent task. When the participant had learnt the task, he or she saw and executed the set of 27 experimental instructions. Along each instruction presentation, participants had the LEGO^{TM} pieces to one side and could see but not touch them.

Type of secondary task was counterbalanced, and blocked for each subject. When the block began, the first screen indicated the type of secondary task. Within each block of secondary task, presentation format was counterbalanced. As a result, participants saw a block of nine items with a particular concurrent task, consisting of three sets according to presentation format (two of the instructions had three-steps, and one had five-steps). Each instruction had the following sequence: after a 500-ms fixation cross the first step was presented; the participant self-administered the steps pressing the space bar, until the last step was shown, followed by the word 'Assemble'. When the participant finished assembling the object, the next instruction was self-administered by pressing the space bar. A second research assistant also registered responses (sequence and position errors).

RESULTS

The effects of format (text, picture, multimedia), concurrent task (spatial, verbal, or no-task), and complexity (three vs. five steps), on assembling accuracy (error rate during assembly) and instruction study times (mean study time per screen) were analyzed using repeated measures ANOVA. Post-hoc paired comparisons were used for significant effects and interactions.

Assembly errors analysis

A repeated measures ANOVA on mean error rate per condition revealed significant effects of format, $F\,(2, 214) = 25.05$,

$p < .001$, $\eta_p^2 = .19$; concurrent task, F (2, 214) = 16.13, $p < .001$, $\eta_p^2 = .13$; and complexity, F (1, 107) = 149.09, $p < .001$, $\eta_p^2 = .58$. Also significant were the interactions between format and concurrent task, F (4, 432) = 6.25, $p < .001$, $\eta_p^2 = .05$; complexity by concurrent task, F (2, 216) = 5.47, $p < .001$, $\eta_p^2 = .05$; and format by complexity, F (2, 216) = 4.65, $p = .013$, $\eta_p^2 = .04$. Table 1 shows means and standard deviations in error rates as a function of format, concurrent task, and complexity.

Regarding the format effect, both multimedia and picture-only conditions led to significantly fewer errors than text-only, and, multimedia obtained fewer errors than picture-only (multimedia $M = .18$, $SD = .15$; picture $M = .21$, $SD = .13$; text $M = .28$, $SD = .16$). As for the concurrent task effect, similar post-hoc paired analyses showed that a concurrent verbal task led to more errors ($M = .26$, $SD = .14$) than a concurrent spatial task ($M = .23$, $SD = .16$), or without a concurrent task ($M = .19$, $SD = .14$); these latter two did not differ significantly. Finally, for the complexity effect, there were significantly less errors in the three-step condition ($M = .16$, $SD = .10$) relative to the five-steps ($M = .29$, $SD = .16$).

Paired post-hoc analyses of the interactions were conducted with Bonferroni correction because of the number of contrasts ($p < .005$). Analyses of the interaction between format and concurrent task showed that mean error rate for the picture-only format was higher with a verbal concurrent task ($M = .29$, $SD = .21$), than with a spatial concurrent task ($M = .19$, $SD = .16$), or without a secondary task ($M = .17$, $SD = .18$), whereas the latter two did not differ. For the text-only format, both secondary tasks led to more errors (verbal secondary task $M = .32$, $SD = .22$; spatial secondary task $M = .29$, $SD = .22$) as compared to no-task ($M = .22$, $SD = .19$). The multimedia condition did not show significant differences as a function of the secondary task. Regarding the format by complexity interaction, in the three-step condition, multimedia presentations led to fewer errors ($M = .12$, $SD = .12$) than both picture-only ($M = .17$, $SD = .12$) and text-only ($M = .19$, $SD = .15$), which did not differ between them. In five-step instructions, multimedia ($M = .24$, $SD = .21$) and picture-only ($M = .26$, $SD = .19$) did not differ between them and were better than text-only ($M = .36$, $SD = .22$). Finally, the interaction between complexity and concurrent task meant that in the three-step presentations fewer errors were observed without a secondary task ($M = .13$, $SD = .11$) compared to both a spatial concurrent task ($M = .17$, $SD = .14$) and a verbal concurrent task ($M = .18$, $SD = .13$). However, in five-step instructions, a verbal secondary task interfered more with performance ($M = .35$, $SD = .21$) than a spatial secondary task ($M = .26$, $SD = .20$) or no-task ($M = .25$, $SD = .20$).

Table 1. Descriptive statistics (mean, SD) of assembly accuracy (mean percent errors) as a function of format, concurrent task and complexity

Condition	Spatial		Verbal		No-task	
	M	SD	M	SD	M	SD
Text, 3	.22	.23	.21	.23	.14	.15
Text, 5	.34	.31	.43	.34	.29	.32
Picture, 3	.16	.16	.20	.22	.14	.18
Picture, 5	.21	.28	.37	.29	.19	.28
Multimedia, 3	.13	.21	.11	.14	.09	.15
Multimedia, 5	.21	.28	.23	.29	.26	.28

3: three-steps
5: five-steps

To sum up, results for errors have shown significant main effects of format, complexity, and concurrent task, as well as significant interactions between format by complexity, format by concurrent task, and complexity by concurrent task. Regarding format, multimedia instructions led to fewer errors than picture-only, whereas text-only format had the most errors. Also, more complex instructions (five vs. three steps) led to more errors. As for secondary task, under a concurrent verbal task participants had more errors than both a spatial secondary task or in single task conditions; these latter two did not differ. The interactions between factors showed that the verbal secondary task produced more errors in both text-only and picture-only instructions (but not multimedia), and also provoked more errors in both complexity levels. The spatial secondary task did not show a pattern of consistent significant effects across format or complexity. As for the format by complexity interaction, both levels of complexity were easier with multimedia and harder with text-only presentations; as complexity increased performance with picture-only instructions rose to the multimedia level. Overall, assembly performance was affected negatively by text-only presentations. Picture-only instructions led to better performance than the text-only ones, were useful as complexity increased, but were susceptible to verbal interferences. Multimedia presentations were generally more efficient, and more resistant to interference. Also, a verbal secondary task led to more errors across different presentation formats.

Study times analysis

One participant had missing times and was eliminated for the following analyses. A repeated measures ANOVA of mean study time per screen as dependent variable revealed significant effects of format, F (2, 214) = 263.68, $p < .001$, $\eta_p^2 = .71$; concurrent task, F (2, 214) = 52.20, $p < .001$, $\eta_p^2 = .33$; and complexity, F (1, 107) = 6.95, $p = .010$, $\eta_p^2 = .06$. Also significant were the interactions between format and concurrent task, F (4, 428) = 15.13, $p < .001$, $\eta_p^2 = .12$; concurrent task and complexity, F (2, 214) = 7.05, $p = .001$, $\eta_p^2 = .06$; and the interaction between all conditions, F (4, 428) = 8.28, $p < .001$, $\eta_p^2 = .07$. Table 2 shows means and standard deviations in study times as a function of format, concurrent task, and complexity.

Regarding the format effect, the text-only condition led to significantly longer study times ($M = 7717$, $SD = 2426$) than picture-only ($M = 4716$, $SD = 2204$) and multimedia ($M = 4854$, $SD = 2072$) conditions, but the latter two did not differ. On the other hand, a concurrent verbal task led to faster study ($M = 5156$, $SD = 1888$) than a concurrent spatial task ($M = 6058$, $SD = 2291$), and no-task ($M = 6073$, $SD = 2222$). As for the complexity effect, participants spent more time per screen in the three-step condition ($M = 5863$, $SE = 2094$) relative to the five-steps ($M = 5661$, $SE = 2091$).

Table 2. Descriptive statistics (mean, SD) of study time (mean per screen) as a function of format, concurrent task, and complexity

Condition	Spatial		Verbal		No-task	
	M	SD	M	SD	M	SD
Text, 3	8827.88	3621.49	6653.33	2321.02	7858.82	2615.19
Text, 5	7801.79	2823.43	7153.64	2899.66	8119.11	4244.28
Picture, 3	5417.28	2980.09	4776.18	2384.54	4471.99	2459.74
Picture, 5	4423.26	2112.85	4394.74	2627.22	4829.29	2638.72
Multimedia, 3	4953.03	2306.89	4307.48	1963.19	5742.93	2976.99
Multimedia, 5	5072.35	2427.39	3773.84	1854.92	5544.71	2743.74

3: three-steps
5: five-steps

Paired post-hoc analyses of the interactions were conducted with Bonferroni correction because of the number of contrasts ($p < .005$). For the interaction between format and concurrent task, study times for the picture-only format did not vary as a function of secondary task (picture-no task: M = 4464.87, SD = 2391.54; picture-spatial: M = 4910.59, SD = 2344.49; picture-verbal: M = 4572.21, SD = 2287.65). In the text-only format, the verbal secondary task led to faster study times (M = 6892.60, SD = 2381.12) compared to no secondary task (M = 7961.42, SD = 2926.71) and spatial secondary task (M = 8296.03, SD = 2913.77). These latter two conditions did not differ. In a similar way, the multimedia format was read slower when presented alone (M = 5592.40, SD = 2599.91) compared with verbal (M = 4003.59, SD = 1805.37) and spatial (M = 4966.70, SD = 2243.54) concurrent tasks. In this case, the verbal secondary task led to faster study than the spatial one.

Similar paired analyses of the interaction between complexity and concurrent task showed that the three-step instructions were read faster when performed with a verbal concurrent task (M = 5218.91, SD = 1908.69), relative to no secondary task (M = 6004.35, SD = 2263.43) or spatial secondary conditions (M = 6367.42, SD = 2604.68); whereas the spatial concurrent task did not significantly differ from the single task conditions. With five-step instructions, a verbal concurrent task also sped up study time, and a spatial secondary task also led to significantly shorter study time than without concurrent task (no secondary task: M = 6141.45, SD = 2561.42; verbal secondary task: M = 5093.35, SD = 2038.87; spatial secondary task: M = 5748.12, SD = 2367.07).

To sum up, results for study times have shown significant main effects of format, complexity, and concurrent task, as well as significant interactions between format by concurrent task, and complexity by concurrent task. Regarding format, text-only instructions led to longer study times. More complex instructions (five vs. three steps) took shorter mean time per screen (corrected for number of steps). As for secondary task, under a concurrent verbal task, participants were faster. The interaction between factors showed that the verbal secondary task accelerated study times when the instructions included words (text or multimedia formats), and also accelerated study across both complexity levels; when complexity increased (five-steps) a spatial secondary task also sped up processing. Overall, study times were always longer with a text-only presentation, but within presentation modalities, a verbal secondary task led to accelerated processing compared with single tasks. In the same vein, more complex instructions led to faster processing, and also both secondary tasks accelerated processing time.

DISCUSSION

Summary of results

Regarding accuracy in assembly, as predicted, we found significantly better performance in multimedia presentations, followed by pictorial instructions; the worst assembly performance was observed in text-only format. As for WM load, we found the predicted effects of concurrent tasks and complexity: dual task requirements and increased number of steps led to more errors. Dual task affected more picture-only and text-only presentations than multimedia instructions. In text- or picture-only presentations, a verbal secondary task caused more errors, whereas a spatial secondary task interfered with text-only instructions. Neither secondary task affected multimedia accuracy.

Study time, as expected, varied as a function of format presentation, concurrent task, and number of steps. Regarding format, participants took longer to process text-only presentations. Regarding load, participants were faster in dual tasks (compared to single task), and in more complex instructions (more steps). A further inspection showed that a verbal secondary task led to faster processing in text-only and multimedia formats, whereas spatial interferences were seen in multimedia.

Is there a multimedia effect?

Multimedia presentations were the most efficient way to convey instructions: they led to fewer errors and were less affected by secondary tasks. This pattern of results is in line with previous findings about the multimedia effect in comprehension and learning from text (e.g. Mayer, 2001, 2014; Mayer & Sims, 1994; Schnotz, 2014), and in particular in procedural texts (Brunyé et al., 2006; Gyselinck et al., 2008; Michas & Berry, 2000; Novick & Morse, 2000; van Genuchten et al., 2014; Zacks & Tversky, 2003).

As for the question of whether picture-only presentations are similar to multimedia (Brunyé et al., 2006; 2007), we found that pictorial conditions had similar processing times but significantly more errors in assembly than multimedia, and were disrupted by verbal interferences, a cross-modal effect. This pattern of results would suggest that the internal representation built in instructions is not solely image-based, and that verbal coding is also required. Thus, the multimedia advantage would reside in its dual coding (e.g. Mayer, 2001, 2014), and not only in the beneficial properties of images.

How is working memory involved in understanding instructions?

Construction and maintenance of an internal model derived from instructions load WM; we had predicted an increase in study times as a function of dual task and complexity. However, we found that in dual task conditions participants accelerated their processing, and had more errors when presentations were text-only or picture-only. Also, regarding complexity, participants spent less time per step, and had more errors, in more complex instructions (five steps). One possible explanation for these unexpected results could be that under dual task requirements in text or picture-only conditions, or when presentations were more complex, WM capacity would be overtaxed. Given WM limits in capacity and duration, longer times per step could mean losing details or sequence. Participants could try to speed their study times to maintain the model's details and procedural sequence and reach as soon as possible the assembly phase. This would result in poorer performance in assembly. A similar speeded viewing time was obtained by Engle et al. (1992). They used a self-paced moving window version to examine whether high and low span participants adopted different strategies in WM tasks. As evidenced in Figures 2 and 4 (last panels, p. 979 and p. 986), mean viewing time for the to-be-remembered items increased linearly with memory load up to two or three items (for low and high WM span, respectively). However, viewing time for the to-be-remembered items declined abruptly for both groups when memory load was 5. In sum, they found that all participants decreased their viewing times with higher load in WM tasks. Similarly, our paradigm required to maintain in WM sequentially presented material. As long as WM capacity was not exceeded, participants would allot enough viewing time to each step, but when the to-be-remembered material increased in length and complexity beyond a certain point, it might have overtaxed WM, leading to speeded viewing times to arrive at the execution phase. In general, these results would suggest that text or picture-only format, long and sequential presentations are difficult to maintain and integrate, overtaxing WM (see also Mayer, 2014).

Are modality specific WM components differentially involved? As for visuo-spatial WM, previous similar studies employing dual task paradigms found contrasting results. Brunyé et al. (2006) obtained specific modality dual task interferences in memory for instructions (spatial interfered with pictorial format), whereas van Genuchten et al. (2014) found that a spatial interference hindered verbally presented performance, but multimedia instructions were not affected. In the present experiment, the spatial secondary task affected accuracy of text-only presentations, as in van Genuchten et al. (2014). Also, the spatial secondary task interfered with study time in multimedia presentations (but did not obtain more errors), and with study time in more complex presentations. Although the spatial concurrent task did not affect all conditions, it could be argued that the visuo-spatial WM was involved in our procedural assembly task, but not overtaxed except when image generation was required, or the participant had to actively integrate picture and words, or when the load increased. This possible explanation was suggested by previous authors (Brunyé et al., 2006; Gyselinck et al., 2002; van Genuchten et al., 2014): images present information in a parallel, integrated fashion, illustrating spatial relations more directly, and therefore when the instructions are pictorial, they do not deplete visuo-spatial WM capacity. In demanding tasks, for example, when generating images from verbal input (van Genuchten et al., 2014) or maintaining longer sequences of images before completion (this experiment), effects of a modality specific dual task are noticed. On the other hand, the differential involvement of visuo-spatial WM could be a function of study goals: because participants were told they would be assembling the objects after learning, they may have strategically engaged more mental imagery (and visuo-spatial WM) than if they expected to only be tested with a simple memory task (as in Brunyé et al., 2006).

As for the possible role of verbal WM, our results indicate that this system would be implicated in understanding instructions, as evidenced by the overall interference caused by a concurrent verbal task on performance, in terms of speeded processing and more assembly errors. This interference was observed not only in text-only instructions, but also in picture-only presentations. This contrasts with Brunyé et al. (2006) who found modality specific interferences (verbal with textual information). Such an overall pattern would point to a relevant role of verbal WM when building a mental model of instructions, and also when executing them. It would seem that, independently of presentation format, verbal coding is needed. In line with Michas and Berry (2000) and Brunyé et al. (2006), we consider that this effect may be attributed to the importance of the verbal code in organizing a procedural situation. Michas and Berry (2000) argued that the verbal code in procedural instructions contains action information, describing the activities that have to be undertaken in each step. Also, verbal codes could have a role in labeling and identifying unfamiliar objects, as suggested in expository text research (Mayer, 2001).

Multimedia presentations were the least affected by dual task requirements: processing times were faster, but accuracy did not suffer. This result is in line with other studies employing similar dual task paradigms (Brunyé et al., 2006; van Genuchten et al., 2014).

CONCLUSIONS, LIMITATIONS, AND PRACTICAL IMPLICATIONS

In conclusion, multimedia instructions enhanced both processing and execution of procedures, thus indicating the importance of adding both a pictorial code and written assembly instructions. Our results indicate that visuo-spatial and verbal WM systems were used to process and execute these instructions and that, in consequence, whenever one of these memory systems was overtaxed, participants showed difficulties, as indicated by accelerated study time and more errors in assembly. Other studies emphasized the role of visuo-spatial WM (Brunyé et al., 2006; van Genuchten et al., 2014); our experiment highlighted the relevance of verbal WM, regardless of the format of the incoming stimuli. Whereas the pictorial format may be more suitable than the textual one to depict visual objects and spatial layouts, a verbal mental code could be needed when comprehending parts and sequences, and executing actions. In particular, verbal coding could be useful to follow a sequence in time, and to identify objects.

As for the limitations of the present study, we have not assessed performance on the secondary task, following a standard procedure, but it would be helpful to collect a concurrent task performance measure. Also, more difficult secondary tasks, or tasks that tap the central executive component, could better characterize the pattern of interferences. In addition, we have manipulated presentation format, but we cannot be sure of what were the participants looking at in multimedia instructions. A future study with think-aloud or eye-tracking methodologies could address this issue. A further issue arises with the fact that multimedia instructions duplicated the same information, whereas in real instructions the information conveyed is usually complementary. Other studies could focus on how information is integrated across modalities. Finally, a limitation resides in the fact that instructions were delivered in sequential and static representations. Other studies have focused on simultaneous presentations (e.g. Brunyé et al., 2007; Michas & Berry, 2000). On the other hand, video or animations, which were expensive and sophisticated some time ago (e.g. Michas & Berry, 2000), are becoming much more common and are a target for future research.

These experimental results provide information on how to construct good instructions for everyday activities, such as assembly brochures for home furniture, appliances, toys, electronic devices, and the like. Instructions benefit from multimedia, and the pictorial and textual information should be clear, short, and not overtax WM. In addition to the relevance of images, our experiment has also found an overall involvement of verbal WM in comprehension of instructions. Therefore, we recommend that in procedural presentations design, special attention could be directed to textual information. Maybe adding information about object identities, sequence order, or any useful tip about performance might act as an organizer. Future experiments can explorer this issue. These recommendations follow and expand the literature on multimedia presentations.

REFERENCES

Aaronson, D., & Ferres, S. (1984). The word-by-word reading paradigm: An experimental and theoretical approach. In D. Kieras, & M. Just (Eds.), *New methods in reading comprehension research* (edn, pp. 31 – 68). Hillsdale, NJ: Lawrence Erlbaum.

Baddeley, A. D. (1986). *Working memory*. Oxford: Oxford University Press.

Baddeley, A. D. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences*, *4*(11), 417–423. DOI:10.1016/S1364-6613(00)01538-2.

Baddeley, A. D. (2002). Is working memory still working? *European Psychologist*, *7*, 85–97. DOI:10.1027//1016-9040.7.2.85.

Baddeley, A. D., & Lieberman, K. (1980). Spatial working memory. In R. S. Nickerson (Ed.), *Attention and performance VIII* (edn, pp. 521 – 539). Hillsdale, N.J.: Erlbaum.

Brunken, R., Plass, J. L., & Leutner, D. (2003). Direct measurement of cognitive load in multimedia learning. *Educational Psychologist*, *38*, 53–61. DOI:10.1207/S15326985EP3801_7.

Brunyé, T. T., Taylor, H. A., & Rapp, D. N. (2007). Repetition and dual coding in procedural multimedia presentations. *Applied Cognitive Psychology*, *22*(7), 877–895. DOI:10.1002/acp.1396.

Brunyé, T. T., Taylor, H. A., Rapp, D. N., & Spiro, A. B. (2006). Learning procedures: The role of working memory in multimedia learning experiences. *Applied Cognitive Psychology*, *20*, 917–940. DOI:10.1002/acp.1236.

Diehl, V., & Bergfeld Mills, C. (2002). Procedural text structure and reader perceptions and performance. *Journal of General Psychology*, *129*(1), 18–35. DOI:10.1080/00221300209602030.

Engle, R. W., Cantor, J., & Carullo, J. (1992). Individual differences in working memory and comprehension: A test of four hypotheses. *Journal of Experimental Psychology Learning Memory and Cognition*, *18*, 972–992. DOI:10.1037/0278-7393.18.5.972.

Friedman, N. P., & Miyake, A. (2004). The relations among inhibition and interference cognitive functions: A latent variable analysis. *Journal of Experimental Psychology: General*, *133*, 101–135.

Geiger, F. J., & Millis, K. K. (2004). Assessing the impact of reading goals and text structures on comprehension. *Reading Psychology*, *25*, 93–110. DOI:10.1080/02702710490435637.

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. DOI:10.1002/acp.823.

Gyselinck, V., Jamet, E., & Dubois, V. (2008). The role of working memory components in multimedia comprehension. *Applied Cognitive Psychology*, *22*, 353–374. DOI:10.1002/acp.1411.

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review*, *87*, 329–354. DOI:10.1037/0033-295X.87.4.329.

Just, M. A., Carpenter, P. A., & Woolley, J. D. (1982). Paradigms and processes in reading comprehension. *Journal of Experimental Psychology: General*, *111*, 228–238. DOI:10.1037/0096-3445.111.2.228.

Marcus, N., Cooper, M., & Sweller, J. (1996). Understanding instructions. *Journal of Educational Psychology*, *88*, 49–63. DOI:10.1037//0022-0663.88.1.49.

Mayer, R. E. (2001). *Multimedia learning*. New York, NY: Cambridge University Press. doi: 10.1017/CBO9781139164603

Mayer, R. E. (2014). Cognitive theory of multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (43-71). New York: Cambridge University Press. doi: 10.1017/CBO9781139547369

Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist*, *38*, 43–52. DOI:10.1207/S15326985EP3801_6.

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. DOI:10.1037//0022-0663.86.3.389.

Michas, I., & Berry, D. (2000). Learning a procedural task: Effectiveness of multimedia presentations. *Applied Cognitive Psychology*, *14*(6), 555–575.

Mills, B. C., Diehl, V. A., Birkmire, D. P., & Mou, L.-C. (1995). Reading procedural texts: Effects of purpose for reading and predictions of reading comprehension models. *Discourse Processes*, *20*, 79–107. DOI:10.1080/01638539509544932.

Murray, D. J. (1968). Articulation and acoustic confusability in short-term memory. *Journal of Experimental Psychology*, *78*, 679–684. DOI:10.1037/h0026641.

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. DOI:10.3758/BF03211824.

Paivio, A. (1971). *Imagery and verbal processes*. New York: Holt, Rinehart, and Winston.

Paivio, A. (1986). *Mental representations: A dual coding approach*. Oxford, England: Oxford University Press. doi: 10.1093/acprof:oso/9780195066661.001.0001

Schnotz, W. (2014). Integrated model of text and picture comprehension. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (72-103). New York: Cambridge University Press. doi: 10.1017/CBO9781139547369.006

Sweller, J., van Merriënboer, J. J. G., & Paas, F. (1998). Cognitive architecture and instructional design. *Educational Psychology Review*, *10*, 251–296. DOI:10.1023/a:1022193728205.

Van Genuchten, E., Van Hooijdonk, C., Schüler, A., & Scheiter, K. (2014). The role of working memory when "Learning How" with multimedia learning material. *Applied Cognitive Psychology*, *28*, 327–335. DOI:10.1002/acp.2998.

Zacks, J. M., & Tversky, B. (2003). Structuring information interfaces for procedural learning. *Journal of Experimental Psychology: Learning, Memory & Cognition*, *20*, 920–933. DOI:10.1037/1076-898X.9.2.88.

