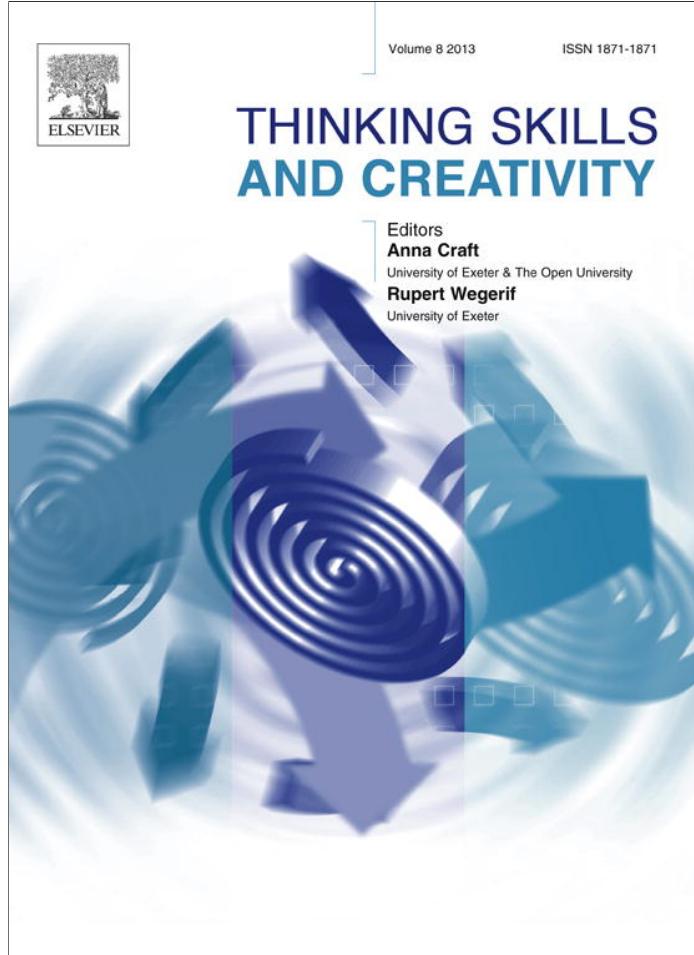


Provided for non-commercial research and education use.  
Not for reproduction, distribution or commercial use.



This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

<http://www.elsevier.com/copyright>



Contents lists available at SciVerse ScienceDirect

# Thinking Skills and Creativity

journal homepage: <http://www.elsevier.com/locate/tsc>

## An examination of the effects of argument mapping on students' memory and comprehension performance

Christopher P. Dwyer\*, Michael J. Hogan, Ian Stewart

School of Psychology, NUI, Galway, Ireland



### ARTICLE INFO

#### Article history:

Received 4 February 2012

Received in revised form 7 November 2012

Accepted 14 December 2012

Available online 31 December 2012

#### Keywords:

Argument mapping

Cognitive load

Immediate recall

Delayed recall

Comprehension

### ABSTRACT

Argument mapping (AM) is a method of visually diagramming arguments to allow for easy comprehension of core statements and relations. A series of three experiments compared argument map reading and construction with hierarchical outlining, text summarisation, and text reading as learning methods by examining subsequent memory and comprehension performance. Effects of study environment, argument size, learning strategy (active and passive) and recall interval (immediate and delayed) were also examined. Results revealed that argument map reading and construction significantly increased subsequent immediate recall for arguments in both passive and active learning settings. These findings indicate that AM is a useful learning and teaching methodology, particularly in comparison with standard text-based learning. Results are discussed in light of research and theory on learning and memory.

© 2013 Elsevier Ltd. All rights reserved.

### 1. Introduction

As much of the knowledge to be acquired by students in school and university requires reading academic textbooks, an important goal for teachers is to aid students in their acquisition of textbook knowledge. However, for long pieces of text, learning can be difficult because the creation of an integrated representation in long-term memory is constrained by ongoing storage limitations in working memory (Cowan, 2000; Miller, 1956). Some researchers have suggested that because it is too memory intensive to remember everything from a passage of text, a macrostructure, or the 'gist' of the text, is stored in long-term memory, and this represents the summary information a reader considers important (Kintsch & van Dijk, 1978). Hence, it is this macrostructure, and not the original text that the reader remembers when later asked to recall the text (Kintsch & van Dijk, 1978). The problem with this learning strategy is that although the formulation of a macrostructure presumably facilitates recall of information, it is likely that information is not encoded at a very deep level of specificity; in other words, the detail of propositions and of relations between propositions will probably not be remembered.

Various organisational strategies have been devised to enhance long-term retention of information, including, for example, summarisation (Kintsch & van Dijk, 1978), hierarchical summarisation (Taylor, 1982) and graphic organisation (Robinson & Kiewra, 1995). Research suggests that when to-be-remembered information is presented in a well-organised manner, the level of free recall is better than when it is presented in a random order (Bower, Clark, Lesgold, & Winzenz, 1969; Myers, 1974). Also, readers who are sensitive to text structure recall more information than readers who are not (Meyer, Brandt, & Bluth, 1980).

\* Corresponding author. Tel.: +353 87 6471823.

E-mail address: [christopher.dwyer.phd@gmail.com](mailto:christopher.dwyer.phd@gmail.com) (C.P. Dwyer).

Hierarchical summarisation is an explicit, active organisational strategy. It involves extracting and summarising the key themes and sub-themes in a text. Taylor (1982) found that the use of hierarchical summarisation (more commonly known as outlining), increased recall of text in students who were trained in the use of the technique. A similar study by Berkowitz (1986) provides a rare example of how organisational strategies can be used to facilitate learning of prose arguments. Berkowitz taught students to construct maps of prose passages. Using this mapping strategy, the main ideas from the passages were summarised in separate boxes and supporting claims were listed as bullet points beneath each of the main ideas. The boxes were organised in a radial structure (i.e. around a central claim). Berkowitz found that for students who used this technique overall recall of passages was significantly improved relative to students who used traditional study techniques (i.e. question-answering and re-reading procedures).

Although Berkowitz described her maps as a graphic representation of the superordinate and some of the more important subordinate ideas in a passage, organised in a manner similar to the way the author organised them in the original selection, the propositional content of the radial maps did not represent fully planned arguments. Also, although Berkowitz attempted to construct maps that corresponded to the way the author organised ideas in the original selection, the radial structures in no way reflected the structure of the argument (see Twardy, 2004 for a discussion of the text to argument map translation process). Therefore, a critical question is whether or not the reading and construction of more explicit, complete, logical, hierarchically structured maps that faithfully represent the structure of an argument can be used as part of a package of classroom learning activities to facilitate students' assimilation of and memory for arguments. One such organisational strategy developed quite recently and that shows particular promise in this regard is argument mapping (e.g. van Gelder, 2000, 2007).

Argument mapping (AM) is a method of visually diagramming arguments in an organised hierarchy, in order to simplify the reading of an argument structure and allow for easy assimilation of core propositions and relations. The AM uses a 'box-and-arrow' design in which boxes represent propositions within an argument while arrows make the inferential relationships between these propositions explicit (see Fig. 1 for an example). Boxes are colour-coded to indicate the nature of propositions (e.g. reasons, objections, and rebuttals), and arrows are labelled so as to specify the nature of the relationship between the propositions (e.g. *but*, *because* or *however*). The use of coloured boxes, arrows connecting boxes, and semantic cues describing relations between propositions are all designed to 'glue' the structure of the argument together and allow the reader to analyse and evaluate a line of reasoning with ease. This system of representation may therefore help to reduce the burden associated with analysing and evaluating text-based argument structures and facilitate subsequent memory and comprehension.

More specifically, when it comes to analysing arguments, the problem with traditional text-based learning is that it does not allow one to readily connect statements that support and dispute specific reasons. The learner must engage in a cognitively demanding process of linking propositions that are located in different paragraphs, on different pages, and so on. When reading a text-based argument, the reader must mentally construct the argument, thus switching attention away from the information presented in the text. In a series of seminal studies, Pollock, Chandler, and Sweller (2002) found that learning is impeded when instructional materials require a high degree of attention switching, for example, between text and figures. They concluded, more generally, that encoding environments that increase the cognitive load placed on the reader tend not only to slow the learning process, but also reduce overall levels of learning.

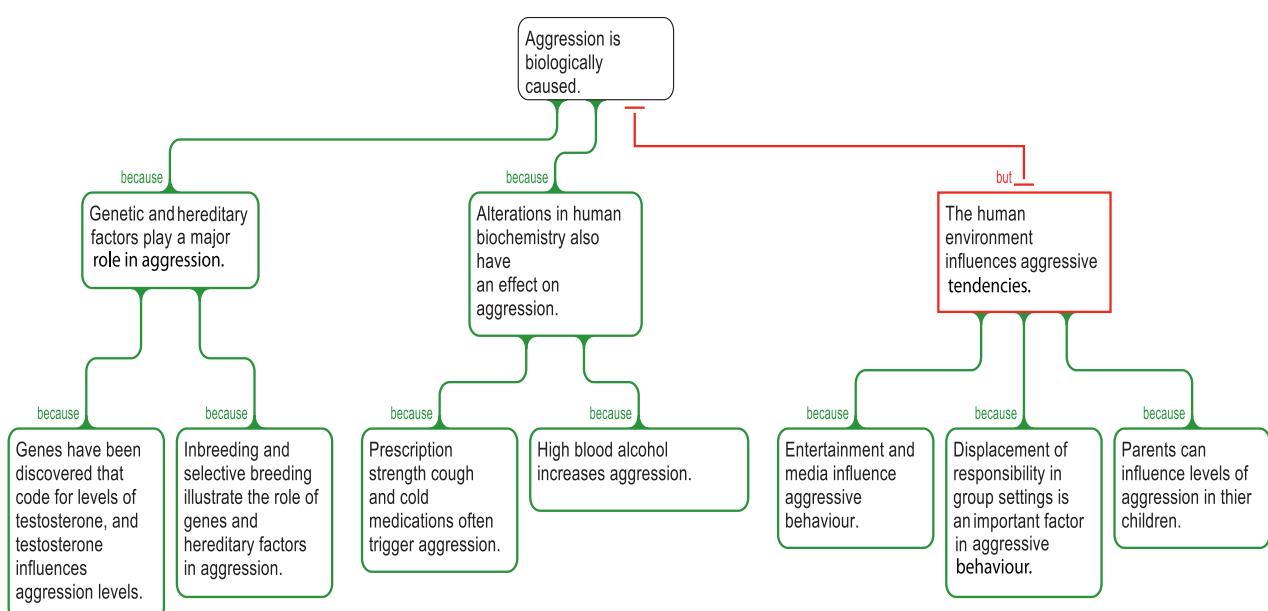


Fig. 1. An example of an argument map built using Rationale™.

The provision of a visual representation of an argument structure may reduce the cognitive load associated with jointly reading and mentally representing the structure of an argument (Dwyer, Hogan, & Stewart, 2010; van Gelder, 2000, 2003). First, AMs utilise Gestalt grouping principles; and research suggests that when to-be-remembered items are organised according to Gestalt principles, such as proximity and similarity, they are better recalled (Jiang, Olson, & Chun, 2000; Woodman, Vecera, & Luck, 2003). For example, AMs place related propositions close to one another, thus complying with the Gestalt grouping principle of proximity. Furthermore, AM uses a consistent colour scheme to highlight reasons, objections, and rebuttals in an argument structure, thus complying with the Gestalt grouping principle of similarity.

Second, unlike standard text, AMs represent arguments through dual modalities (i.e. both verbal and visual–spatial forms of representation), thus facilitating both the latent information processing capacity of individual learners and better recall beyond that of encoding information associated with either verbal or visual–spatial aspects alone (Mayer, 1997; Paivio, 1986, 1971). Dual-coding theory and research (Paivio, 1971, 1986), Mayer's (1997) conceptualisation and empirical analysis of multimedia learning, and Sweller and colleagues' research on cognitive load (Sweller, 2010), suggests that learning can be enhanced and cognitive load decreased by the presentation of information in a visual–verbal dual-modality format (e.g. diagram and text), provided that both visual and verbal forms of representation are adequately integrated (i.e. to avoid attention-switching demands). Given that AMs support dual-coding of information in working memory via integration of text into a diagrammatic representation, cognitive resources previously devoted to translating prose-based arguments into a coherent, organised and integrated representation are 'freed up' and can be used to facilitate deeper encoding of arguments in AMs, which in turn facilitates successful retrieval (Dwyer et al., 2010). Thus, it may be that AM can provide a system of argument representation that may help readers to better encode, assimilate and recall arguments.

Third, AM presents information in a hierarchical manner. When arguing from a central claim, one may present any number of argument levels which need to be adequately represented for the argument to be properly conveyed. For example, an argument that provides a (1) support for a (2) support for a (3) support for a (4) claim has four levels in its hierarchical structure. More complex or 'deeper' arguments (e.g. with three or more argument levels beneath a central claim) are difficult to represent in text due to its linear nature; and yet it is essential that these complex argument structures are understood by a student if their goal is to remember the argument (Pollock et al., 2002). However, AMs make explicit the hierarchical structure of the argument, thus alleviating working memory of cognitive load associated with assimilating the structure of the argument, as would be the case in studying traditional text.

Although argument maps have been in existence for almost 200 years (Buckingham-Shum, 2003; Reed, Walton, & Macagno, 2007; van Gelder, 2007), it is only with the advent of various argument mapping software programmes, such as Rationale™ (van Gelder, 2007), that the time required to construct an argument map has been substantially reduced, as the construction of an argument using this software needs only the choice of an appropriate box and associated relational cue, the typing of text into the box, and the selection of an appropriate location for the box in the argument structure (i.e. in relation to other propositions).

Though other forms of argument diagramming exist, such as concept mapping and mind-mapping (Buzan & Buzan, 1997), they differ from argument mapping based on the manner in which they are organised and the way in which each 'proposition' is presented. For example, the problem with many concept mapping techniques is that they do not present an argument *per se*. Instead, they present a graphical structure that acts as a representation of a separate text, which might be used to diagram: the links among concepts, decision-making schemes, a set of plans or instructions, or at best, act as an argument overview – which does not represent the argument in full. Thus, because the text of the argument and the diagram may often be separate entities, concept mapping may become more cognitively demanding by adding the necessity of switching attention from text to diagram and vice versa (e.g. Chandler & Sweller, 1991; Pollock et al., 2002; Tindall-Ford, Chandler, & Sweller, 1997). In addition, if the reader of a concept map is not familiar with the information from which the map is derived, then the map itself becomes meaningless. Thus in this context, concept mapping strategies may not necessarily be useful pedagogical aids that are open to analysis by everyone.

A small number of studies suggest that AM may help to improve overall critical thinking after extended deliberate practice (Butchart, Bigelow, Oppy, Korb, & Gold, 2009; van Gelder & Rizzo, 2001; van Gelder, Bissett, & Cumming, 2004) and memory ability after minimal exposure to argument maps (Dwyer et al., 2010). For example, a recent study conducted by Dwyer et al. (2010) suggests that the reading of AMs results in better memory performance than the reading of text. Six groups of students were given 10 minutes to study the argument '*Computers can think*' (adapted from Horn, 1999) and were subsequently tested for both memory and comprehension. Arguments were presented to groups in six different conditions: (1) a 30-proposition (small) text, (2) a 50-proposition (large) text, (3) a 30-proposition monochrome argument map, (4) a 50-proposition monochrome argument map, (5) a 30-proposition colour argument map, or (6) a 50-proposition colour argument map. Results revealed that those who studied from AMs (regardless of colour or size manipulation) scored significantly higher on the memory test than those who studied from text.

The current set of three experiments sought to extend previous research by examining whether the apparently beneficial effects of AM are transferable across (1) different study and testing environments, (2) time (i.e. from immediate memory to delayed memory testing), (3) different study strategies (i.e. passive and active learning) and (4) different topics studied, that is, in comparison with text reading, text summarisation and hierarchical outlining. The core manipulations and hypotheses associated with each experiment are summarised in Table 1.

**Table 1**

The core manipulations and hypotheses associated with each experiment.

Experiment	Design	Hypotheses
<b>1</b> The effects of study format and setting on immediate recall and comprehension	2 (study format) $\times$ 2 (setting) MANCOVA – independent variables (IVs): study format (text, argument map) and setting (group test, individual test). Dependent variables (DVs): comprehension, immediate recall	Argument map group will show better memory and comprehension than text group, regardless of environmental context. No effect on memory or comprehension due to setting
<b>2</b> The effects of study format and argument size on immediate and delayed recall	2 $\times$ 2 $\times$ 2 mixed ANCOVA. IVs: format (text and map) and argument size (30 and 50 propositions). DVs: immediate and delayed recall	Argument map group will show better immediate and delayed recall than text group. Beneficial effects of argument map condition will be greater for larger argument structures. Better immediate recall performance when compared with delayed recall assessment
<b>3</b> The effects of different active learning strategies on immediate recall	Between Subjects ANCOVA. IV: learning strategy (3 levels: summarisation, outlining, argument mapping). DV: immediate recall	Argument map translation exercise fosters better immediate recall than outlining or summarisation. Outlining exercise fosters better recall memory than summarisation

## 2. Method

### 2.1. Overall procedure

Three experiments employing first year psychology undergraduates as participants were conducted over a 2-year period. In each experiment, an argument was presented to students for purposes of study and afterward a cued recall test was administered. The latter consisted of a set of statements from the original study materials. Underneath each statement was either 'Because', or 'But', followed by a set of blank lines. Participants were asked to fill in the correct reason or objection to each statement based on the argument previously studied. Memory tests were scored by two independent raters who used a standardised scoring manual to code responses. Experiment 1 (but not Experiments 2 or 3) also included a comprehension test, which consisted of a set of 12 of the propositions from the original study material. Each question asked whether the statement was either a *reason* or an *objection* to the central claim: *Computers can think*.

In all three experiments, all participants were first introduced to concepts of thinking skills, critical thinking and AM in a lecture delivered in the two weeks prior to commencement of the experimental work. In addition, the verbal and spatial sub-scales of the *Differential Aptitude Test* (Bennett, Seashore, & Wesman, 1986) were administered. The verbal reasoning sub-test of the DAT tests analogical reasoning by providing sentences with missing words (e.g. "... is to night as breakfast is to ...") and five word-pairs to choose from that complete the sentence correctly (e.g. supper – corner, supper – morning, corner – morning, etc.). The spatial reasoning sub-test asks participants to mentally fold cut-outs to produce an object, the correct one being embedded in an array of four choices. Previous research on the DAT revealed that internal consistency reliability coefficients for verbal reasoning are consistently in the .90s, while those for spatial reasoning are consistently in the .80s (Anastasi, 1988; Wang, 1993). Students were provided with 20 min to complete both reasoning tests, each of which consisted of 20 items.

Given that the reading and construction of AMs requires the ability to reason both verbally and spatially, both verbal and spatial reasoning ability were assessed as baseline measures and included as covariates in the analysis of group differences, in order to control for any pre-existing differences in ability among groups assigned to different experimental conditions. In addition, as described in Baddeley's (2000, 2002) model of working memory, Paivio's (1986) dual-coding theory and the general framework for thinking, memory can be seen as consisting of both a visual–spatial coding system and a verbal coding system, which may work collaboratively or independent of one another during perceptual and encoding processes. As a result, it is possible that students with different levels of verbal and spatial reasoning ability would approach the text reading and AM reading tasks in different ways. For example, when studying an AM, a student with higher verbal reasoning ability and lower spatial reasoning ability may focus on the verbal aspects of the map and not make full use of the spatial arrangement of propositions for the purpose of comprehending arguments or committing them to memory. On the other hand, a student with higher spatial reasoning ability and lower verbal reasoning ability may focus on the spatial arrangement of detail on the map and give less attention to the text.

Those with high verbal and spatial ability may have the capacity to focus on both aspects, that is, the verbal detail of the propositions as well as their relational structure in the spatial array of the AM. In this context, given that high levels of both abilities may moderate the relationship between AM and text reading on memory and comprehension outcomes, it is necessary to control for both verbal and spatial reasoning in order to ensure that differences between the AM group and text group are not a result of higher levels of either reasoning ability. It was further hypothesised that higher verbal and spatial reasoning scores would correlate with higher memory and comprehension scores (Hitch & Baddeley, 1976; Jiang et al., 2000; Luck & Vogel, 1997).

## 2.2. Experiment 1

Experiment 1 had two aims: (1) to replicate research conducted by Dwyer et al. (2010), by examining differences between AM reading and text reading using a 30-proposition version of the argument, 'Computers can think' and (2) to extend this previous research by examining whether or not memory and comprehension performance varied as a function of study environment, with students examined in both individual and group settings. Based on research by Dwyer et al. (2010), it was hypothesised that students who use AMs to study would perform better than those who study from text on memory testing. It was further hypothesised that those who study from AMs would perform better on comprehension testing, particularly in the individual study setting. Specifically, because (1) comprehension is a more complex task than memorisation (Anderson & Krathwohl, 2001) and because (2) social presence can negatively affect performance on more complex tasks (Bond & Titus, 1983; Wühr & Huestegge, 2010); it was hypothesised that those in the isolated study group would perform better than those in the group study setting on comprehension.

Research suggests the environment in which an individual studies information may affect how the information is encoded and ultimately how it is retrieved (Godden & Baddeley, 1975; Smith & Vela, 2001; Tulving, 1984). Given that one important goal of the current set of experiments was to examine the utility of AMs as a pedagogical aid, it was important to examine the difference between classroom use of AMs and the use of AMs as individual study aids. In the current experiment, it was hypothesised that the beneficial effects of AMs would be greater in an individual (i.e. isolated study) setting when compared with a group study setting as potentially less cognitive load would be imposed on students in the individual study condition (e.g. social distractions that may cause attention switching; Pollock et al., 2002; Tindall-Ford et al., 1997). Specifically, Baron (1986) suggests that the presence of others may act as distracting stimuli, which may cause the switching of attention from the intended focus of attention. Research by Mullen (1987) also suggests that group settings can distract an individual's focus away from the task at hand by focusing attention towards the self (e.g. *How do I look? Or Is my behaviour acceptable?*). Furthermore, though past research has suggested that completion of cognitive tasks in the presence of others can increase motivation (or *drive*) of those completing the task (Zajonc, 1965, 1980) and attentional focus (Klauer, Herfordt, & Voss, 2008), a meta-analysis by Bond and Titus (1983) found that the presence of others can have a negative effect on performance for more complex tasks. Consistent with the above hypothesis, Bond and Titus' findings are also supported by more recent research conducted by Wühr and Huestegge (2010), who found that social presence can negatively affect performance on cognitive tasks due to the cognitive load (Sweller, 2010) associated with competition for working memory resources and/or attentional resources.

### 2.2.1. Participants

Participants were first year arts students ( $N=131$ ; 93 females, 38 males), aged between 18 and 25 years. In return for their participation students were awarded academic course credits.

### 2.2.2. Materials and measures

Experimental materials were constructed by extracting a sub-set of the arguments presented in Robert Horn's (1999) AM: *Can computers think?* Two sets of study materials for this argument were developed: a 30-proposition text and a 30-proposition argument map.

Memory was assessed using a 'fill-in-the-blanks' cued recall test that assessed memory for reasons and objections linked to each of the major arguments supporting or refuting the central claim: *Computers can think*. The test consisted of 14 questions and was scored on a scale of 0–28. Using the intra-class correlation coefficient (ICC) method, the inter-rater reliability (as measured by two raters) for the memory test was .90. Comprehension was scored on a scale of 0–12; students were asked to decide if a sub-set of 12 of the original propositions (present in both study conditions) either supported or refuted the central claim: *Computers can think*.

### 2.2.3. Procedure

Participants were assigned to study in either a lecture hall (in the presence of other students completing the same tasks) or alone in an isolated booth. After the study environment was allocated, participants in each study environment were then randomly assigned to study the topic 'Computers can think' from either a piece of text or from an AM. All participants were allotted 10 min to read their assigned study materials and were instructed to learn the material with a view to being tested. After 10 min had elapsed, study materials were collected and the cued recall test was administered. After 10 min, the memory assessments were collected and the comprehension test was administered. Five minutes were provided for completion after which tests sheets were collected and participants were debriefed and thanked.

### 2.2.4. Results

Means and standard deviations are presented in Table 2. A  $2 \times 2$  between subjects MANCOVA was used to assess the effects of study materials and environment on comprehension and recall of arguments, while controlling for baseline verbal and spatial reasoning ability. The two between-subjects factors were study materials (text versus argument) and environment (individual versus group setting).

A significant main effect of study materials ( $F[1,125]=11.68$ ,  $MSE=13.71$ ,  $p=.001$ , partial  $\eta^2=.09$ ), was found with AM participants showing higher recall than text participants. There was also a significant main effect of study environment

**Table 2**

Descriptive statistics for study material × environmental context on immediate recall &amp; comprehension.

Condition	M	Std. deviation	N
Memory			
Lecture hall setting			
Text	5.80	3.87	39
Argument Map	7.97	4.79	37
Isolated setting			
Text	4.29	2.64	27
Argument Map	6.18	3.81	28
Comprehension			
Lecture hall setting			
Text	7.08	2.25	39
Argument map	6.58	2.21	37
Isolated setting			
Text	7.44	2.64	27
Argument Map	6.54	2.63	28

( $F[1,125] = 7.04$ ,  $MSE = 13.71$ ,  $p = .009$ , partial  $\eta^2 = .05$ ), with group setting participants showing higher recall than isolated setting participants. There was no study material × study environment interaction effect on memory.

There was no significant effect of study material or study environment or any interaction effect on comprehension performance. Both memory and comprehension were significantly correlated with verbal reasoning (Memory:  $r = .33$ ,  $p = .015$ ; Comprehension:  $r = .45$ ,  $p = .001$ ). However, neither was correlated with spatial reasoning ability.

### 2.2.5. Discussion of Experiment 1

Results from Experiment 1 revealed that those who studied from AMs recalled more information than those who studied from text, regardless of the environmental setting, which is consistent with research by Dwyer et al. (2010). The results also revealed that, contrary to the proposed hypothesis, students who studied and were tested in a lecture hall setting performed better than those who studied and were tested in an individual study booth. In this experiment, participants in the lecture hall setting were monitored to ensure that they did not consult or copy from one another. In the absence of consultation or copying, it may be that the lecture hall setting provided a more formal and stimulating testing venue. It is also possible that students recognised the presence of others as a potential distractor, thus leading to added attentional focusing; in which students' ability to screen out distractions was enhanced (Klauer et al., 2008). Alternatively, it is further possible that participants may have felt more *drive* and responded more competitively in a group setting – in other words, the effect may have been due to some sort of social facilitation (Zajonc, 1965, 1980), which facilitated memory performance even in the context of what was assumed to be a relatively complex argument reading task.

While a group setting may affect participants in particular ways that an isolated setting does not (e.g. distraction, attention switching and social facilitation), such effects were equal across experimental manipulations of materials in the current study. Also, post hoc analysis revealed that, in comparison with text reading, AM reading was associated with better subsequent memory performance in the group setting and the isolated test setting ( $p < .05$  for both comparisons). As such, the beneficial effects of AM reading on memory performance were robust in both an isolated test setting and a group setting, notwithstanding the potential impact of social distraction, social facilitation and other factors linked to group study and testing. In any event, one purpose of this series of experiments was to elucidate the effect of different study techniques in the context of the classroom; and thus, the large group setting has ecological and face validity. As such, all subsequent experiments were conducted in a lecture hall setting only.

Results of the current experiment also revealed that there was an effect of neither study environment, nor AM reading on comprehension performance; a finding which is again consistent with research by Dwyer et al. (2010). Mean scores on the comprehension test suggest that students were performing on average not very far above chance levels (they scored between 6.5 and 7.5 out of 12 on average in a two-choice proposition classification test). In other words, it may be that students did not deeply comprehend the argument and when they were unsure as to whether or not a proposition either supported or refuted the central claim, they simply guessed; and as there were only two answers to choose from, they could guess the correct answer approximately half the time. At the same time, it is notable that students who scored higher on the comprehension test also performed better on the verbal reasoning sub-test of the DAT. Those high on verbal reasoning ability may have engaged in some form of higher-order metacognitive thinking (e.g. critical thinking) while working to learn from the study materials. Thus, while the findings suggest that the reading of AMs facilitates memory for arguments, it may be that some additional training in AM is needed to foster the kinds of metacognitive skills that may be necessary to support deeper comprehension.

Research suggests that to perform optimally in tests of comprehension and memory, one should read first to understand and then re-read in order to memorise (Pollock et al., 2002). However, it is possible that because students were aware that they would be tested after they studied, they assimilated information strictly for purposes of memorisation and sacrificed the need to truly comprehend the argument structure. Thus, it may be that there was not enough time to study materials for

purposes of truly evaluating and comprehending the argument, at least, not enough time to distinguish text reading from AM reading in this context.

Nevertheless, in the presence of a significant correlation between verbal reasoning and memory performance, and controlling for this co-variation, there remained a significant effect of stimulus materials on memory performance. In other words, although verbal reasoning ability also predicted memory performance, unlike for comprehension test performance, the reading of AMs provided some additional benefits in terms of memory for individual propositions in a cued-recall test. It may also be that a certain amount of information can be remembered in the absence of truly understanding the relational structure of an argument, and that AMs facilitate memory for discrete propositions in this context by highlighting each proposition in a distinct box.

Given the apparent memorial advantage of AM over text as shown in this first experiment, one further question that immediately arises is whether or not this benefit would be sustained over time; that is, from immediate recall (as assessed in this experiment) to delayed recall (i.e. one week later). Retention of information over longer durations is of course important; not alone for the purpose of acquiring knowledge, but also for deepening comprehension. For example, it may be the case that the deeper comprehension of arguments is not only dependent on extensive reading and learning, but also on sufficient construction of schemas, which are built in order to retain information for relatively long periods of time. This schema-building process may allow for connections between arguments encoded at various times to be integrated in different configurations and at different levels of complexity. In order to investigate whether advantages of AM reading versus text reading would be sustained over time, Experiment 2 sought to examine the impact of study materials on immediate and delayed recall for both smaller and larger argument structures.

### 2.3. Experiment 2

Experiment 2 had three key aims. The first was to replicate research conducted by Dwyer et al. (2010), by examining differences between small and large study materials – whether or not the beneficial effects of AM on memory extend to larger materials that impose a greater amount of cognitive load. Though Dwyer and colleagues previously found a larger benefit for smaller argument structures over larger argument structures, it may have been that the effects of argument size they observed were a function of the topic being studied. Dwyer and colleagues had speculated that the beneficial effects of AM would be greater for larger argument structures because graphical representation would be increasingly beneficial as cognitive load increases (Chandler & Sweller, 1991; Sweller, 1999, 2010). However, first year arts students may have been largely unfamiliar with the arguments presented in the 'Computers can think' argument structure, and thus AM reading in this experiment may have evoked a slower, more analytical reading of arguments which moderated additional beneficial effects of AM reading over text reading as argument size, increased. Thus, the second aim of Experiment 2 was to make comparisons between AM and text reading, but in the context of a different study topic. In research by Dwyer et al. (2010) and the research conducted in Experiment 1, students studied and were tested on the topic *Computers Can Think*. In Experiment 2, students studied and were tested on the topic *Aggression is Biologically Caused*. This was done in order to investigate whether the effects seen in previous experiments would be replicated in the context of a different topic.

The final aim of this experiment was to extend previous research on AM's effect on recall by examining whether or not the beneficial effects of AM extend across time. Thus, students' memory was tested both immediately after studying and again a week later (i.e. delayed recall). It was hypothesised that beneficial effects of AM reading would be greater for delayed over immediate recall assessment. More specifically, it was hypothesised that benefits of organisation and dual-coding (i.e. visual and verbal coding) on memory (Mayer, 1997; Paivio, 1986, 1971) are sustained for longer, thus the forgetting rate from immediate to delayed recall assessment would be less for AM reading when compared with text reading.

#### 2.3.1. Participants

Two-hundred and eighty-six participants (204 females, 82 males) took part in the initial (study and assessment) session, while, due to attrition in attendance, only 199 (142 females, 57 males) took part in the second (assessment only) session.

#### 2.3.2. Materials and measures

Study materials, each of which presented the topic '*Aggression is biologically caused*' included (1) a 30-proposition (small) text, (2) a 50-proposition (large) text (3) a 30-proposition AM and (4) a 50-proposition AM. Memory was assessed using a 'fill-in-the-blanks' cued recall test with 12 questions that probed for reasons and objections linked to major arguments supporting or refuting the claim '*Aggression is biologically caused*' and was scored on a scale of 0–24. Inter-rater reliability (based on the ICC method) for this test was .95. The delayed recall test was the same as the immediate recall test except for the order of presentation of questions.

#### 2.3.3. Procedure

Participants were randomly allocated to one of four groups in which they studied from either a 30-proposition AM, a 50-proposition AM, a 30-proposition text, or a 50-proposition text. Study materials were distributed and participants were instructed to learn the material with a view to being tested. Participants were provided 15 min to study their materials in this experiment (i.e. 5 min longer than in Experiment 1), in order to provide those learning from both smaller and larger argument structures with enough time to assimilate the information provided. After 15 min had elapsed, study materials

**Table 3**Descriptive statistics for study material  $\times$  size on immediate and delayed recall.

Condition	M	Std. deviation	N
Small			
Immediate recall			
Text	7.76	3.65	63
Argument map	10.83	4.41	65
Delayed recall			
Text	6.28	3.64	50
Argument map	7.58	4.20	52
Large			
Immediate recall			
Text	5.26	3.41	81
Argument map	6.47	3.62	77
Delayed recall			
Text	3.94	3.37	56
Argument map	4.05	2.90	41

were collected and the cued recall test was administered. Participants were given 10 min to complete this test, after which tests were collected and the class concluded. The following week, participants returned and were given a 10 min re-test on the same material.

### 2.3.4. Results

Given the substantial variation in sample size for immediate and delayed recall assessments, three separate ANCOVAs were conducted. A 2 (study materials)  $\times$  2 (size) ANCOVA was first conducted to examine the effects of the study materials and material size on immediate recall, while controlling for baseline verbal and spatial reasoning ability. Results from the ANCOVA revealed a significant main effect of size ( $F[1,280] = 65.67$ ,  $MSE = 12.21$ ,  $p < .001$ , partial  $\eta^2 = .19$ ), with the smaller argument structure group showing significantly better recall performance than the larger argument structure group. Results from the ANCOVA also revealed a significant main effect of study material ( $F[1,280] = 24.17$ ,  $MSE = 12.21$ ,  $p < .001$ , partial  $\eta^2 = .08$ ), with the argument map group showing better recall performance than the text group. However, there was no size  $\times$  study material interaction. Descriptive statistics are shown in Table 3. Recall performance was significantly correlated with both verbal ( $r = .33$ ,  $p < .001$ ) and spatial ( $r = .24$ ,  $p < .001$ ) reasoning ability.

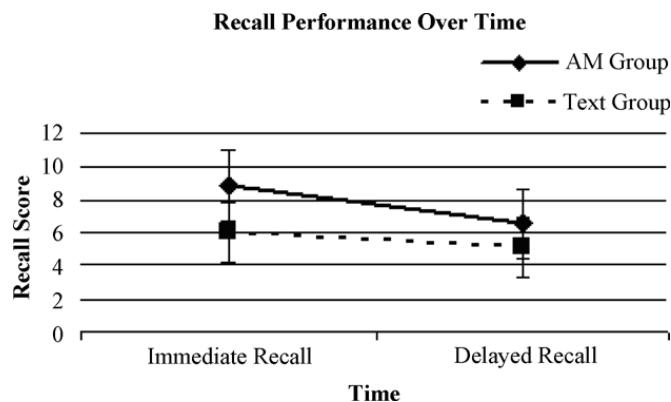
A 2 (study materials)  $\times$  2 (size) ANCOVA was conducted to examine the effects of study materials and size of materials on delayed recall, controlling for baseline verbal and spatial reasoning ability. Results revealed a significant main effect of size ( $F[1,193] = 39.83$ ,  $MSE = 11.05$ ,  $p < .001$ , partial  $\eta^2 = .17$ ), with the smaller argument structure group showing higher delayed recall performance than the larger argument structure group. There was no main effect of study material, nor any size  $\times$  study material interaction. Delayed recall was significantly correlated with both verbal ( $r = .34$ ,  $p < .001$ ) and spatial ( $r = .20$ ,  $p = .004$ ) reasoning ability.

A 2 (study materials)  $\times$  2 (size)  $\times$  2 (testing times) mixed ANCOVA was conducted to examine memory performance differences between study groups at two testing times, both immediately after the study period and one week later. There was a main effect of study materials ( $F[1,193] = 8.29$ ,  $MSE = 20.31$ ,  $p = .004$ , partial  $\eta^2 = .04$ ), with the argument map group scoring significantly higher than the text group. There was also a main effect of size ( $F[1,193] = 51.46$ ,  $MSE = 20.31$ ,  $p < .001$ , partial  $\eta^2 = .21$ ), with the smaller argument structure group scoring higher than the larger argument structure group. Though there was no main effect for time, there was a significant time  $\times$  study material interaction ( $F[1,193] = 10.46$ ,  $MSE = 3.49$ ,  $p = .001$ , partial  $\eta^2 = .05$ ), whereby the benefits of argument map over text were greater for immediate recall relative to delayed recall (see Fig. 2). There was no size  $\times$  time interaction, nor was there a three way interaction.

### 2.3.5. Discussion of Experiment 2

The findings from Experiment 2 further confirm that AMs can facilitate better immediate recall for arguments than text reading. In addition, this finding suggests that argument map superiority is not dependent on topic studied, as similar results were found in Experiment 1, which used a different study topic. This experiment also examined delayed recall. Overall memory performance decreased on average from testing time 1 to testing time 2. In addition, against our initial hypothesis, the superiority of AM seen for immediate recall did not show strong transfer to delayed recall. Although the AM group scored higher on average than the text group on delayed recall, the former also showed a greater decrease in performance from immediate to delayed recall than the latter.

Results revealed that large argument groups showed poorer recall than small argument groups. These data indicate a threshold in terms of number of propositions that can be reasonably assimilated in a short space of time, regardless of format or time of assessment. For example, recall of 12 target memories from a set of 50, after a study period of 15 min, is more difficult than being asked, given a similar study period, to recall the same 12 target memories from a set of 30. This is consistent with cognitive load theory (Pollock et al., 2002), which suggests that study environments that overburden memory are associated with poor learning outcomes (Sweller, 1999).



**Fig. 2.** Memory performance of those in both study material groups from test time 1 to time 2.

Findings from this experiment also appeared to confirm the pattern of findings from the Experiment 1, as there were correlations between verbal reasoning and both immediate recall performance; providing further evidence to suggest that those who are proficient at verbal reasoning are able to recall more information from their studying, even when it has been a week since they studied that information. In addition, unlike Experiment 1, there was a significant correlation between spatial reasoning and immediate recall, as well as with delayed recall.

Though Experiments 1 and 2 yielded some very interesting findings, in order to comprehensively evaluate AM as a learning tool, it must be assessed not only as a passive learning aid, but also as an active learning aid. This line of research is reasonable to consider, given that AM was designed specifically as a means of actively structuring arguments (van Gelder, 2007), rather than as a method of passively assimilating arguments or presenting them to others. The importance of this active construction of arguments is also relevant given recent research findings regarding the beneficial effects on learning outcomes of infusing active learning into classroom settings (e.g. Burbach, Matkin, & Fritz, 2004; Butchart et al., 2009; Perry, Huss, McAuliff, & Galas, 1996). As AM software is designed specifically for active construction of arguments, a logical progression in AM research is to assess whether active learning conducted through AM construction enhances memory performance. Thus, while Experiments 1 and 2 assessed the effects of passively studying from previously constructed AMs on students' recall performance, Experiment 3 examined whether active learning through AM construction enhanced students' recall performance.

#### 2.4. Experiment 3

The aim of Experiment 3 was to compare the effects of active AM, hierarchical outlining and standard text summarisation on immediate recall performance in the context of the active learning and study of arguments. Previous research has demonstrated the beneficial effects of active study on learning outcomes (e.g. Burbach et al., 2004; Hake, 1998; Laws, Sokoloff, & Thornton, 1999; Perry et al., 1996; Redish, Saul, & Steinberg, 1997). In addition, all three strategies examined in Experiment 3 can be usefully employed to actively study and assimilate text-based arguments (Butchart et al., 2009; Kintsch & van Dijk, 1978; Taylor, 1982; Taylor & Beach, 1984). However, it can be argued that AMs offer particular advantages in this regard, given that the active construction of AMs may make the relationships among propositions clearer via its 'box-and-arrow' format, thus enhancing the ability to assimilate arguments (van Gelder, 2003). Therefore, it was hypothesised that students who actively construct and study AMs would perform better on immediate recall assessment than those who actively construct and study either outlines or text summaries. It was also hypothesised that students who actively construct and learn from hierarchical outlines would perform better on immediate recall assessment than those who actively learn using text summarisation techniques. The latter prediction is based on the facts that hierarchical outlines, like AMs, present information in a hierarchically structured manner and that they have been shown to significantly enhance learning in previous research (Robinson & Kiewra, 1995; Taylor, 1982; Taylor & Beach, 1984).

##### 2.4.1. Participants

Participants were first year Arts students ( $N=136$ ; 97 females, 39 males), aged between 18 and 25 years. In return for their participation students were awarded academic course credits.

##### 2.4.2. Materials and measures

All groups were provided with a common study text which contained 30 propositions which argued for and against the claim: *'There was nothing wrong with the United States Supreme Court's decision to uphold the individual's right to bear arms'*. Participants were also provided with materials appropriate to their particular condition: Argument Mapping: a 30-proposition AM with 12 boxes left blank; Hierarchical Outlining: a 30-proposition outline with 12 proposition lines left blank; Text Summarisation: a blank page. The texts provided to those in all three conditions were identical, but study tools (i.e. AM, Outline, and Text summarisation) were different for each condition.

Memory was assessed using a 'fill-in-the-blanks' cued recall test that assessed memory for reasons and objections linked to each of the major arguments supporting or refuting the central claim: '*There was nothing wrong with the United States Supreme Court's decision to uphold the individual's right to bear arms*'. The test consisted of 12 items and was scored on a scale from 0 to 24. Inter-rater reliability (based on ICC) of the test was .92.

#### 2.4.3. Procedure

Participants were randomly allocated to one of three groups in which they studied via one of three active learning techniques: text summarisation, hierarchical outlining or AM. In all three groups, participants were given a 30-proposition text to read, based on the central claim '*There was nothing wrong with the United States Supreme Court's decision to uphold the individual's right to bear arms*.' and were asked to transfer a portion of the information (i.e. 12 propositions) into another document using a particular technique. Participants in the AM condition were given an incomplete AM and were asked to "complete the argument map by filling in the 12 blank boxes with the correct propositions from the associated text". Participants in the outlining group were given an incomplete hierarchical outline and were asked to "complete the outline by filling in the 12 blank spaces with the correct propositions from the associated text". In the text summarisation group (i.e. the control group), participants were given a blank sheet of paper and were instructed to "take notes as you would naturally, summarising the text on the blank sheet of paper provided, with a specific focus on the last paragraph of the text" (i.e. the same portion of text that was subject to transfer in the other conditions).

Participants were given five minutes to read the text (which contained 30 propositions) and a further 20 min to actively learn using their designated method. After the 25 min had elapsed, study materials were collected. The cued recall test was then administered with ten minutes allotted for completion. Finally, the completed recall tests were collected and participants were debriefed and thanked. Only those who took part in the active learning portion of the study session (i.e. specifically, those who filled in at least four boxes on the AM, four blank lines on the outline or wrote down four propositions on the blank page) were included in the data analysis. Taking into account the amount of time allowed to complete the task, it was decided that completion of a third or more of the allotted exercise indicated a genuine effort to both read and actively learn.

#### 2.4.4. Results

A between-subjects ANCOVA was used to analyse the effects of the three active learning techniques on immediate recall. The between-subjects factor was study technique (text summarisation, outlining and AM) while the covariates were baseline verbal and spatial reasoning ability. Preliminary analysis evaluating homogeneity-of-slopes indicated that in the case of both verbal and spatial reasoning, immediate recall did not differ significantly as a function of experimental condition. Results from the ANCOVA revealed a significant main effect of study technique ( $F[2,131] = 16.35$ ,  $MSE = 11.09$ ,  $p < .001$ , partial  $\eta^2 = .20$ ), with the AM group scoring significantly higher on memory testing than the text summarisation group; and with the outline group also showing significantly higher memory performance than the text summarisation group ( $F[1,131] = 30.68$ ,  $MSE = 11.09$ ,  $p < .001$ ,  $d = .84$ ). Descriptive statistics are shown in Table 4. No difference was found between the AM and outline groups on memory performance. Results further revealed that memory performance was significantly correlated with verbal reasoning ( $r = .40$ ,  $p < .001$ ), but not with spatial reasoning ( $r = .14$ ,  $p = .098$ ).

#### 2.4.5. Discussion of Experiment 3

The findings from Experiment 3 showed that both AM and hierarchical outline construction facilitated better immediate recall performance than text summarisation. One reason for the apparent superiority of both AM and outline construction over text summarisation is that in the two former methods, information must be hierarchically organised, which research suggests can improve memory performance (Robinson & Kiewra, 1995; Taylor, 1982; Taylor & Beach, 1984), while in the latter, though participants may have organised information hierarchically, they need not have. For example, participants in the text summarisation condition may have elected to use a summarisation method that did not employ hierarchical organisation thus putting them at a potential disadvantage compared with those in the other groups.

It was hypothesised that the spatial organisation of propositions within an AM might facilitate superior recall in comparison with hierarchical outlines. Though the AM group did show better recall on average than the outlining group, the difference between the two groups was non-significant. This pattern suggests that though the spatial organisation of propositions in an AM may benefit memory, the hierarchical organisation of propositions may be the most critical factor influencing subsequent recall.

Finally, consistent with the results from previous experiments, a correlation was found between verbal reasoning and immediate recall. This finding provides further confirmation of the link between verbal reasoning ability and memory for

**Table 4**

Descriptive statistics for effects of actively learning from study materials on immediate recall.

Condition	M	Std. deviation	N
Text summarisation	4.84	3.25	37
Outlining	7.76	3.68	51
Argument mapping	9.42	3.77	48

arguments in classroom learning contexts – a relationship that appears to hold both for more passive reading and more active learning conditions. However, as in Experiment 1, there was no correlation between recall and spatial reasoning ability, thus suggesting that spatial reasoning ability is not a critical factor determining one's ability to recall specific propositions in an argument, regardless of whether or not AMs were used to study the argument.

### 3. General discussion

#### 3.1. Interpretation of results

Results from this series of experiments suggest that learning through AM reading and construction facilitates subsequent recall of propositions in an argument structure and that AM study seems to be superior in this respect to a number of more traditional study formats, including passive text-reading and active text summarisation. In addition, the superiority of AM reading and construction in this respect appears to hold across study settings (i.e. isolated study settings and lecture hall settings) and topics studied. Nevertheless, as observed in Experiment 2, AM's ability to facilitate enhanced recall over traditional text-based reading was not well maintained over time. In Experiment 3, results revealed that those who actively learned via constructing an AM or a HO performed significantly better than those in the text summarisation group on the recall assessment, most likely as a result of the explicit hierarchical organisation of information in both the AM and HO conditions (Dwyer et al., 2010; Taylor, 1982; Taylor & Beach, 1984).

As previously discussed, a growing body of literature suggests that organisational strategies can be used to facilitate recall performance and learning in general (e.g. Berkowitz, 1986; Dwyer et al., 2010; Myers, 1974; Meyer et al., 1980; Oliver, 2009; Robinson & Kiewra, 1995; Taylor, 1982; Taylor & Beach, 1984). Importantly, the organisational structure of text may not always facilitate learning and memory. Text is sequential, but the argument presented within text is not and thus, the nature of text does not allow one to link propositions that support or dispute specific claims (van Gelder, 2003). Overall, we attribute AM's superiority over text-based studying to the hierarchical structuring of information in AMs. Notably, any group condition which studied from hierarchically organised study materials (e.g. AM and HO) recalled significantly more than groups which did not have hierarchically organised materials. This finding is consistent with previous research by Taylor (1982) and Taylor and Beach (1984) who found that the use of hierarchical summarisation increased recall and comprehension performance of those who used the technique.

Results from Experiment 2 showed that those who studied small argument structures, regardless of format, scored significantly higher on both immediate and delayed recall performance than those who studied from large study materials. Given the increased cognitive load associated with larger arguments, this finding was not surprising as research suggests that study environments that overburden memory are associated with poor learning outcomes (Sweller, 1999). However, it was also found that AM seemed to provide less of an advantage over other formats in the context of the larger (50 proposition) map than in the context of the smaller (30 proposition) map. This finding is converse to our stated hypothesis that the superiority of AM studying would be greater when cognitive load is increased, as AMs are designed partly to decrease cognitive load. Overall, the data suggest that there is a threshold in terms of the number of propositions that can be reasonably assimilated in a short space of time, regardless of study format. This finding supports previous research regarding the limited capacity of working memory and how an increase in the amount of information that is necessary to assimilate contributes to cognitive load (e.g. Cowan, 2000; Pollock et al., 2002; Sweller, 1999, 2010).

The correlation between memory performance and both verbal and spatial reasoning ability was also measured in all the experiments. In every case, verbal reasoning was significantly correlated with memory performance, which is consistent with findings reported by Dwyer et al. (2010). It is hypothesised that verbal reasoning and recall were correlated because those who are proficient at verbal reasoning are able to retain information long enough in order to analyse and evaluate propositions and their relations in an argument structure, which in turn facilitates their ability to answer subsequent test questions correctly (Colom, Abad, Rebollo, & Shih, 2005; Hitch & Baddeley, 1976). This interpretation is also consistent with the results of Experiment 1, in which verbal reasoning was also found to be correlated with comprehension performance.

Spatial reasoning was only found to be correlated with recall test performance in Experiment 2. This finding suggests that spatial reasoning ability can influence memory for arguments in a text or an AM in specific learning contexts, but that verbal reasoning ability may be a more robust and consistent predictor of memory in educational contexts. Nevertheless, the prime reason that the spatial reasoning assessment was administered in this research was in order to control for baseline ability. For example, those better at spatial reasoning, may be better suited to study from AMs than from text.

#### 3.2. Limitations and future research

Though the results of these experiments are informative in terms of the comparison between AM and alternative learning techniques, there are some limitations to the research that must be considered. One possible limitation is that there are student characteristics other than verbal and spatial reasoning ability that may have influenced learning and may have been usefully included in the analysis, such as motivation or interest in study topics and learning methods, for example. In retrospect, it is reasonable to assume those with higher levels of motivation to learn or to out-perform others, or those who found certain study methods novel and interesting (e.g. argument mapping or active learning), may have been more motivated to succeed and thus, may have outperformed others not similarly motivated or interested (Marzano, 1998; Pintrich,

2000; Pintrich & DeGroot, 1990). For example, the lack of an effect of AM reading on comprehension in Experiment 1 may have been caused by a lack of motivation to engage in higher-order thinking processes (e.g. critical thinking) that may be critical drivers of comprehension in novel learning environments. In addition, participants who were interested in new learning techniques – even by the slightest manipulation to the technique, for example, colour in an AM, may have been further motivated by this novelty to perform better. In other words, in addition to a possible main effect of motivation on performance (Garcia, Pintrich, & Paul, 1992; Pintrich, 2000), it is possible that there was an unobserved interaction effect between motivation and novelty of learning materials across experiments. Having some knowledge of students' motivation may also have shed light on the attrition of students from testing time 1 to 2 in Experiment 2. Thus, the differential impact of instructional techniques on participants who were highly motivated should have been compared with the effects on those participants who were poorly motivated.

Another possible limitation that should be considered is that methods for gauging the difficulty of the topic studied, or students' familiarity with or interest in the topic, were not used. It may have been that memory performance was not a result of recall ability alone, but could also be a result of interest or familiarity with the topic studied (Taylor & Beach, 1984). Three different topics were used in this set of experiments and yielded similar patterns of findings. Nevertheless, participant ratings of familiarity, difficulty or intrinsic interest could yield potentially useful, additional information; perhaps especially with respect to unexpected patterns of findings such as the lack of an effect for comprehension in Experiment 1 or idiosyncratic patterns such as that seen for spatial reasoning in Experiment 2. Notably, though student's motivation, interest and familiarity with the topics studied in Experiments 1–3 would all have been useful, one general challenge in these experiments was the issue of working within the allotted class time of 50 min. This time constraint made the inclusion of multiple assessments difficult for the efficient running of experiments with large numbers of students. Nevertheless, future research on AM's effect on key aspects of performance, such as memory and comprehension, needs to address some of the limitations discussed thus far, including motivation to learn, familiarity with the topic studied, perceived difficulty of the topic studied and interest in the topic studied.

This research aimed to compare AM and traditional presentation formats. However, one difficulty in making a fair comparison between these formats is that participants have been exposed to traditional formats such as text reading, all their lives, whereas they were new to AMs. Though the training provided in these experiments (i.e. a 50 min lecture), may have been enough for participants to perform well in terms of studying from and constructing semi-completed AMs, it may not have been enough for the full potential of the advantageous effects of AM reading and construction on delayed recall and comprehension to be realised. If the AM training provided was insufficient with respect to the reading of AMs, then this can be seen as a possible explanation for the null effects on comprehension performance in Experiment 1 and delayed recall performance in Experiment 2.

In order for a more equal comparison to be made, participants would likely require more substantial training in the use of AM than that provided in these experiments. This view is consistent with van Gelder et al.'s (2004) view on the potential beneficial effects of AM on critical thinking; that is, in terms of achieving optimal growth in critical thinking ability, one must also consider what level of 'deliberate practice' is needed. Naturally, it is not simply the method or tools of instruction and learning that are important when working to cultivate critical thinking ability, but also the intensity and quality of practice. The question arises in this context as to what level of AM familiarity and practice is needed to facilitate optimal memory and comprehension performance when using AM reading or construction as a classroom exercise. Guidelines on what constitutes sufficient training in AM, in the context of critical thinking training studies, have been provided by van Gelder et al. (2004) and suggest a semester long course in critical thinking taught through argument mapping for purposes of improving critical thinking ability. Though it was not an aim of this research to examine argument mapping's effect on critical thinking, it is hypothesised, based on van Gelder and colleagues' research, that any semester-long course that teaches a subject through AMs will adequately teach students how to use argument mapping as a study method to facilitate memory and comprehension ability. As a result, an aim for future research should be to examine the effects of explicit training in AM and the related method of hierarchical summarisation (i.e. outlining) on memory and comprehension.

Finally, given that AMs represent arguments through dual modalities (i.e. both verbal and visual-spatial forms of representation), and given that students may potentially prefer a visual-based learning style over a verbal-based learning style or visa versa, future research could compare the differential effects of AM study and construction activities on the performance of students who differ in their visual versus verbal thinking styles. Furthermore, there are a number of learning/thinking styles that can be assessed in relation to effects of AM on memory and comprehension (e.g. judicial thinking, legislative thinking and executive thinking styles; Sternberg, 1999). Thus, future research should examine the effects of AM on memory and comprehension performance of students with varying learning styles.

### 3.3. Summary and conclusion

An important goal for teachers should be to aid students in their acquisition of textbook knowledge. For example, through the teaching of organisational strategies that can help improve their comprehension (Meyer et al., 1980) and memory (Sweller, 1999); with the added goal of keeping cognitive load to a minimum. Based on AM's hypothesised ability to keep constraining factors like cognitive load to a minimum (van Gelder, 2000, 2001), the current research aimed to replicate and extend findings from research conducted by Dwyer et al. (2010) by examining the effects of AM on memory and

comprehension through manipulating study environment (i.e. isolated or group settings), length of time information is retained (i.e. immediate and delayed recall) and learning methods (passive and active learning).

The results of these experiments demonstrated that, when compared with traditional text-based information delivery methods (e.g. linear prose), AM reading and construction significantly increased subsequent immediate recall for arguments in both passive and active learning settings. However, further research examining the effects of AM training on delayed recall and comprehension is also necessary. Overall, the findings of the current research recommend AM as an efficacious methodology for learning and teaching compared to other protocols and particularly in comparison with standard text-based learning.

## Acknowledgement

This research was funded by the Irish Research Council for Humanities and Social Sciences.

## References

Anastasi, A. (1988). *Psychological testing* (6th ed.). New York: MacMillan Publishing.

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

Baddeley, A. D. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences*, 4, 417–423.

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

Baron, R. S. (1986). Distraction-conflict theory: Progress and problems. *Advances in Experimental Social Psychology*, 19, 1–40.

Bond, C. F., Jr., & Titus, L. J. (1983). Social facilitation: A meta-analysis of 241 studies. *Psychological Bulletin*, 94(2), 265–292.

Bennett, G. K., Seashore, H. G., & Wesman, A. G. (1986). *Differential aptitude test, verbal reasoning subtest spatial reasoning subtest* (British ed.). Windsor: Psychological Corporation & NFER-Nelson.

Berkowitz, S. J. (1986). Effects of instruction in text organization on sixth-grade students' memory for expository reading. *Reading Research Quarterly*, 21(2), 161–178.

Bower, G. H., Clark, M. C., Lesgold, A. M., & Winzenz, D. (1969). Hierarchical retrieval schemes in recall of categorised word lists. *Journal of Verbal Learning and Verbal Behavior*, 8, 23–343.

Buckingham-Shum, S. J. (2003). The roots of computer supported argument visualization. In P. A. Kirschner, S. Buckingham-Shum, & C. Carr (Eds.), *Visualizing argumentation: Software tools for collaborative and educational sense-making* (pp. 3–24). London: Springer-Verlag.

Burbach, M., Matkin, G., & Fritz, S. (2004). Teaching critical thinking in an introductory leadership course utilizing active learning strategies: A confirmatory study. *College Student Journal: ProQuest Education Journals*, 38(3), 482.

Butchart, S., Bigelow, J., Oppy, G., Korb, K., & Gold, I. (2009). Improving critical thinking using web-based argument mapping exercises with automated feedback. *Australasian Journal of Educational Technology*, 25(2), 268–291.

Buzan, T., & Buzan, B. (1997). *The mind map book*. London: BBC Books.

Chandler, P., & Sweller, J. (1991). Evidence for cognitive load theory. *Cognition and Instruction*, 8(4), 351–362.

Colom, R., Abad, F. J., Rebollo, I., & Shih, P. C. (2005). Memory span and general intelligence: A latent-variable approach. *Intelligence*, 33, 623–642.

Cowan, N. (2000). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral & Brain Sciences*, 24, 87–185.

Dwyer, C. P., Hogan, M. J., & Stewart, I. (2010). The evaluation of argument mapping as a learning tool: Comparing the effects of map reading versus text reading on comprehension and recall of arguments. *Thinking Skills and Creativity*, 5(1), 16–22.

Garcia, T., Pintrich, P. R., & Paul, R. (1992). Critical thinking and its relationship to motivation, learning strategies and classroom experience. In *Paper presented at the 100th annual meeting of the American Psychological Association* Washington, DC, August 14–18.

Godden, D. R., & Baddeley, A. D. (1975). Context-dependent memory in two natural environments: On land and underwater. *British Journal of Psychology*, 66, 325–331.

Hake, R. (1998). Interactive-engagement vs. traditional methods: A six-thousand student survey of mechanics test data for introductory physics courses. *American Journal of Physics*, 66(1), 64–74.

Hitch, G. J., & Baddeley, A. D. (1976). Verbal reasoning and working memory. *Quarterly Journal of Experimental Psychology*, 28, 603–621.

Horn, R. E. (1999). *Can computers think?* MacroVU, Inc.

Jiang, Y., Olson, I. R., & Chun, M. M. (2000). Organization of visual short-term memory. *Journal of Experimental Psychology: Learning Memory and Cognition*, 26, 683–702.

Kintsch, W., & van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review*, 85, 363–394.

Klauer, K. C., Herfordt, J., & Voss, A. (2008). Social presence on the Stroop task: Boundary conditions and an alternative account. *Journal of Experimental Social Psychology*, 44, 2. <http://dx.doi.org/10.1016/j.jesp.2007.02.009>

Laws, P., Sokoloff, D., & Thornton, R. (1999). Promoting active learning using the results of physics education research. *UniServe Science News*, 13, 14–19.

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, 390, 279–281.

Marzano, R. J. (1998). *A theory-based meta-analysis of research on instruction*. Aurora, CO: Mid-Continent Regional Educational Laboratory. Retrieved from: [http://www.mcrel.org/pdf/instruction/5982rr\\_instructionmeta\\_analysis.pdf](http://www.mcrel.org/pdf/instruction/5982rr_instructionmeta_analysis.pdf)

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

Meyer, B. J. F., Brandt, D. M., & Bluth, G. J. (1980). Use of top-level structure in text: Key for reading comprehension of ninth-grade students. *Reading Research Quarterly*, 16(1), 72–103.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *The Psychological Review*, 63, 814–897.

Mullen, B. (1987). Self-attention theory: The effects of group composition on the individual. In B. Mullen, & G. Goethals (Eds.), *Theories of group behaviour* (pp. 125–146). New York: Springer-Verlag.

Myers, J. L. (1974). *Memory for prose material*. Amherst, MA: University of Massachusetts.

Oliver, K. (2009). An investigation of concept mapping to improve the reading comprehension of science texts. *Journal of Science Education and Technology*, 18(5), 402–414.

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

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

Perry, N. W., Huss, M. T., McAuliff, B. D., & Galas, J. M. (1996). An active-learning approach to teaching the undergraduate psychology and law course. *Teaching of Psychology*, 23(2), 76–81.

Pintrich, P. R. (2000). The role of goal orientation in self-regulated learning. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 451–502). San Diego, CA: Academic Press.

Pintrich, P., & DeGroot, E. (1990). Motivational and self-regulated learning components of classroom academic performance. *Journal of Educational Psychology*, 82, 33–40.

Pollock, E., Chandler, P., & Sweller, J. (2002). Assimilating complex information. *Learning & Instruction*, 12, 61–86.

Reed, C., Walton, D., & Macagno, F. (2007). Argument diagramming in logic, law and artificial intelligence. *The Knowledge Engineering Review*, 22(1), 87–109.

Redish, E., Saul, J., & Steinberg, R. (1997). On the effectiveness of active-engagement microcomputer-based laboratories. *American Journal of Physics*, 65(1), 45.

Robinson, D. H., & Kiewra, K. A. (1995). Visual argument: Graphic organizers are superior to outlines in improving learning from text. *Journal of Educational Psychology*, 87(3), 455–467.

Smith, S. M., & Vela, E. (2001). Environmental context-dependent memory: A review and meta-analysis. *Psychonomic Bulletin & Review*, 8(2), 203–220.

Sternberg, R. J. (1999). *Thinking styles*. New York: Cambridge University Press.

Sweller, J. (1999). Instructional design in technical areas. In *Australian education review No. 43*. Victoria: Acer Press.

Sweller, J. (2010). Cognitive load theory: Recent theoretical advances. In J. L. Plass, R. Moreno, & R. Brünken (Eds.), *Cognitive load theory* (pp. 29–47). New York: Cambridge University Press.

Taylor, B. M. (1982). Text structure and children's comprehension and memory for expository material. *Journal of Educational Psychology*, 74, 323–340.

Taylor, B. M., & Beach, R. W. (1984). The effects of text structure instruction on middle-grade students' comprehension and production of expository text. *Reading Research Quarterly*, 19, 134–146.

Tindall-Ford, S., Chandler, P., & Sweller, J. (1997). When two sensory modes are better than one. *Journal of Experimental Psychology: Applied*, 3(4), 257–287.

Tulving, E. (1984). Precis of elements of episodic memory. *Behavioral and Brain Sciences*, 7, 223–268.

Twardy, C. R. (2004). Argument maps improve critical thinking. *Teaching Philosophy*, 27(2), 95–116.

van Gelder, T. J. (2000). Learning to reason: A reason-Able! Approach. In C. Davis, T. J. van Gelder, & R. Wales (Eds.), *Cognitive Science in Australia, 2000. Proceedings of the fifth Australasian cognitive science society conference Causal*. Adelaide.

van Gelder, T. J. (2001). How to improve critical thinking using education technology. In G. Kennedy, M. Keppell, C. McNaught, & T. Petrovic (Eds.), *Meeting at the crossroads: Proceedings of the 18th annual conference of the Australian Society for Computers in learning in tertiary education* (pp. 539–548). Melbourne: Biomedical Multimedia Unit, University of Melbourne.

van Gelder, T. J. (2003). Enhancing deliberation through computer supported argument mapping. In P. Kirschner, S. Buckingham Shum, & C. Carr (Eds.), *Visualizing argumentation: Software tools for collaborative and educational sense-making* (pp. 97–115). London: Springer-Verlag.

van Gelder, T. J. (2007). The rationale for rationale. *Law, Probability & Risk*, 6, 23–42.

van Gelder, T. J., Bissett, M., & Cumming, G. (2004). Enhancing expertise in informal reasoning. *Canadian Journal of Experimental Psychology*, 58, 142–152.

van Gelder, T. J., & Rizzo, A. (2001). Reasonable across the curriculum in is it an odyssey in learning? In *Proceedings of the 2001 conference of the computing in Education Group of Victoria*.

Wang, L. (1993). *The differential aptitude test: A review and critique*. Austin, TX: Annual Meeting of the Southwest Educational Research Association.

Woodman, G. F., Vecera, S. P., & Luck, S. J. (2003). Perceptual organization influences visual working memory. *Psychonomic Bulletin & Review*, 10(1), 80–87.

Wühr, P., & Huestegge, L. (2010). The impact of social presence on voluntary and involuntary control of spatial attention. *Social Cognition*, 28(2), 145–160.

Zajonc, R. B. (1965). Social facilitation. *Science*, 149(3681), 269–274.

Zajonc, R. B. (1980). Compresence. In P. B. Paulus (Ed.), *Psychology of group influence* (pp. 35–60). Hillsdale, NJ: Erlbaum.