



Representational transformations: Using maps to write essays

Adam Binks^{*a}, Alice Toniolo^a, Miguel A. Nacenta^b

^a School of Computer Science, University of St Andrews, Jack Cole Building, St Andrews, KY16 9SX, UK

^b Department of Computer Science, University of Victoria, 3800 Finnerty Rd, Victoria, BC V8P 5C2, Canada

ARTICLE INFO

Keywords:

Argument mapping
Externalization
Multimedia
Argumentation
Representation
Translation

ABSTRACT

Essay-writing is a complex, cognitively demanding activity. Essay-writers must synthesise source texts and original ideas into a textual essay. Previous work found that writers produce better essays when they create effective intermediate representations. Diagrams, such as concept maps and argument maps, are particularly effective. However, there is insufficient knowledge about how people use these intermediate representations in their essay-writing workflow. Understanding these processes is critical to inform the design of tools to support workflows incorporating intermediate representations.

We present the findings of a study, in which 20 students planned and wrote essays. Participants used a tool that we developed, Write Reason, which combines a free-form mapping interface with an essay-writing interface. This let us observe the types of intermediate representations participants built, and crucially, the process of how they used and moved between them.

The key insight is that much of the important cognitive processing did not happen within a single representation, but instead in the processes that moved between multiple representations. We label these processes 'representational transformations'. Our analysis characterises key properties of these transformations: cardinality, explicitness, and change in representation type. We also discuss research questions surfaced by the focus on transformations, and implications for tool designers.

1. Introduction

External representations are often used to support complex and cognitively demanding activities. For example, someone writing an essay might first create a collection of index cards with their ideas and lay them out on a cork board before typing a draft of the essay on a computer. Similarly, novelists, game designers, and writers of overhead presentations might rely on a variety of external representations such as sketches, whiteboard annotations, or even paper prototypes to carry out cognitively demanding tasks.

Previous research has explored the affordances and advantages of different external representations (such as notes Slotte and Lonka, 2001, sketches Cherubini et al., 2007, concept maps Novak and Cañas, 2006 and essays McGinley, 1992). Different kinds of external representations (hereafter, 'representations') are better at supporting different kinds of tasks (Hegarty, 2004; Lohse et al., 1994; Zhang, 1997; Zhang and Norman, 1994). For example, Zhang and Norman (1994) found that different representations of the Tower of Hanoi problem affected people's ability to solve it. People found solutions most easily when

representations provided useful memory aids and made important task information visually explicit. Software interfaces can also support the agile creation of different types of representations. Many tools are constructed under the premise that a better visual layout and interactive design could facilitate complex cognitive tasks (Badam et al., 2019; Cañas et al., 2004; Carneiro et al., 2019; Introne, 2009; Wang et al., 2019).

Understanding the characteristics of different representations and how people perceive and use them is key for the design of interfaces that effectively support complex tasks. In the domain of essay-writing interfaces, the area we focus on in this research, there is evidence that intermediate representations such as concept maps can be very useful (Fan et al., 2019; Jafari and Zarei, 2015; Slotte and Lonka, 2001; Zarina and Fatima, 2015). However, there is insufficient knowledge about how people use intermediate representations within the workflow of the task. We therefore know little about the role that multiple representations and the interactions between them play in the larger process of essay-writing, which involves additional processes beyond the editing of the text, such as elicitation, evaluation and organisation of ideas.

* Corresponding author.

E-mail address: ab390@st-andrews.ac.uk (A. Binks).

<https://doi.org/10.1016/j.ijhcs.2022.102851>

Received 27 January 2022; Received in revised form 9 May 2022; Accepted 10 May 2022

Available online 13 May 2022

1071-5819/© 2022 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

As a next step towards understanding the role of multiple representations in interfaces, in this paper we focus on the task of persuasive essay-writing because essay-writing is pervasive in educational and professional contexts, important (Prosser and Webb, 1994), and cognitively demanding (Biggs, 1988). We designed Write Reason¹ (Fig. 1) — a hybrid tool with both text and map representations — to study how people use these representations in an essay-writing task.

We ran a pilot study and a main study, in which students planned and wrote essays using Write Reason. Our goal was to understand how different representations are used within the workflow of an essay-writing task. Through a detailed — mostly qualitative — analysis of data collected from the 20 essay writers in our main study², we characterised the types of intermediate representations employed by the participants and observed how they moved between them. A key insight is that the value of representations came not only from the individual representations but from the process by which writers transformed one representation into another. We describe how this relatively neglected concept of representational transformation is crucial and can be influential for the design of new interfaces that support complex cognitive tasks.

Some key findings (described in greater detail in the discussion, Section 7):

- Many participants chose to create multiple intermediate representations, such as additional maps or texts. Participants transformed between representations in a batch (all at once) or interleaved (bit by bit) manner.
- We observed no simple one-to-one transliterations between the map and text representations: participants always added or removed ideas.
- Participants' representations varied substantially: we observed frequent use of arrows to map argumentative structure, but also other approaches like spatial clustering and appropriation of special nodes to denote pros/cons.

In this paper, we first review the existing evidence on representations and transformations. We then describe the design of our hybrid tool, and the results of a preliminary study to validate its experimental value. Next, we present the method and results of our main study. Finally, we discuss these results and their implications for practitioners.

2. Related work

Here, we introduce the domain of persuasive essay-writing, and how essay outcomes are evaluated. We then survey literature on the power of representations and tools supporting them across a range of tasks, including essay-writing. We conclude with existing work on how interaction between representations supports task workflows, arguing that this is a relatively neglected topic with important unanswered questions.

2.1. Persuasive essay-writing

Essay-writing typically involves selecting pieces of information from sources, and finding connections between them (McGinley, 1992; Spivey, 1990). Persuasive essay-writing is a kind of essay-writing where the writer aims to persuade the reader of the truth of some proposition (Kinneavy, 1971). Typically, writers aim to achieve this by constructing a compelling argument, so when reading sources they aim to find *argumentative* connections between claims (Spivey, 1990). To help novice writers achieve these goals, research focuses on determining (1)

¹ Write Reason is available online at <http://adambinks.me/write-reason/editor>, with a video overview of the interface.

² Participants' representations can be explored interactively at <http://adambinks.me/write-reason/explore>.

how to support student's essay-writing skills (Carroll et al., 2016; Newell et al., 2011; Prosser and Webb, 1994; Smith et al., 1999) and (2) how to measure the quality of their essay outcomes (Biggs, 1988; Lavelle and Zuercher, 2001; Reddy and Andrade, 2010)

Hayes and Flower (1987) identify that writers use three major processes: planning, where writers generate and organise ideas; sentence generation, to produce explicit sentences; and revising, where writers evaluate and attempt to improve their work. Using a think-aloud protocol, Hayes and Flower identify that these processes are typically interwoven, often in a recursive way. For example, when revising their work, a writer may decide to add an extra paragraph, and then plan and write it before continuing the revision. The challenge of effectively interweaving these processes to combine and connect ideas from different sources makes essay-writing a cognitively demanding task (Hayes and Flower, 1987; McGinley, 1992).

There are various ways to evaluate essay outcomes. Most commonly, educators create rubrics to align their essay marking criteria with the intended learning outcomes (Reddy and Andrade, 2010). The Structure of Observed Learning Outcome (SOLO) taxonomy is a popular general framework (Chan et al., 2002; Wong, 2007) that grounds the development of specific rubrics for specific learning outcomes. It is often applied to essay-writing (Biggs, 1988; Lavelle and Zuercher, 2001; Prosser and Webb, 1994; Smith et al., 1999). SOLO categorises students' work into five hierarchical levels of depth of understanding. Undergraduate student essays tend to achieve two of the deeper levels of understanding: multi-structural, and relational (Campbell et al., 1998). *Multi-structural* essays are arrangements of sequential points, focusing on knowledge-telling, with only simple arguments at the local level — overarching argumentative structure is largely absent. *Relational* essays are structured as an argument. Different perspectives are compared and synthesised on the whole essay scale, with a clear conceptualisation of the essay as a coherent, unified structure. Relational essays tend to result from a deeper engagement with the ideas, and attain higher grades than multi-structural essays (Prosser and Webb, 1994). These categories are important in our work to examine the effect of representational structures on essay outcomes.

2.2. Representations

We now focus on how using different representations can affect mental processes and outcomes in essay-writing, and in general.

Zhang's theory of *representational determinism* (Zhang, 1997) identifies the power of external representations to affect mental processes and their outcomes. Representational determinism is the view that "the format of a representation can determine:

- (A) What information can be perceived,
- (B) What processes can be activated,
- (C) What structures can be discovered from the specific representation."

For example, presenting people with different visual representations of the Tower of Hanoi problem (Zhang and Norman, 1994) and Tic Tac Toe (Zhang, 1997) affects their ability to solve the puzzles and identify winning strategies because the different representations lead viewers to perceive different information (A) and discover different structures (C). Representational determinism motivates the development and identification of modes of representation which activate helpful cognitive processes (B), and cause viewers to perceive important information (A) and structures (C). In persuasive essay-writing, it is important to perceive relevant claims in sources (A), and discover argumentative structures between them (C) through important reasoning processes such as hypothesising, questioning evidence, and developing ideas and arguments (B) (McGinley, 1992). Therefore, representational determinism implies that identifying representations that activate these reasoning processes and encouraging writers to use them may improve

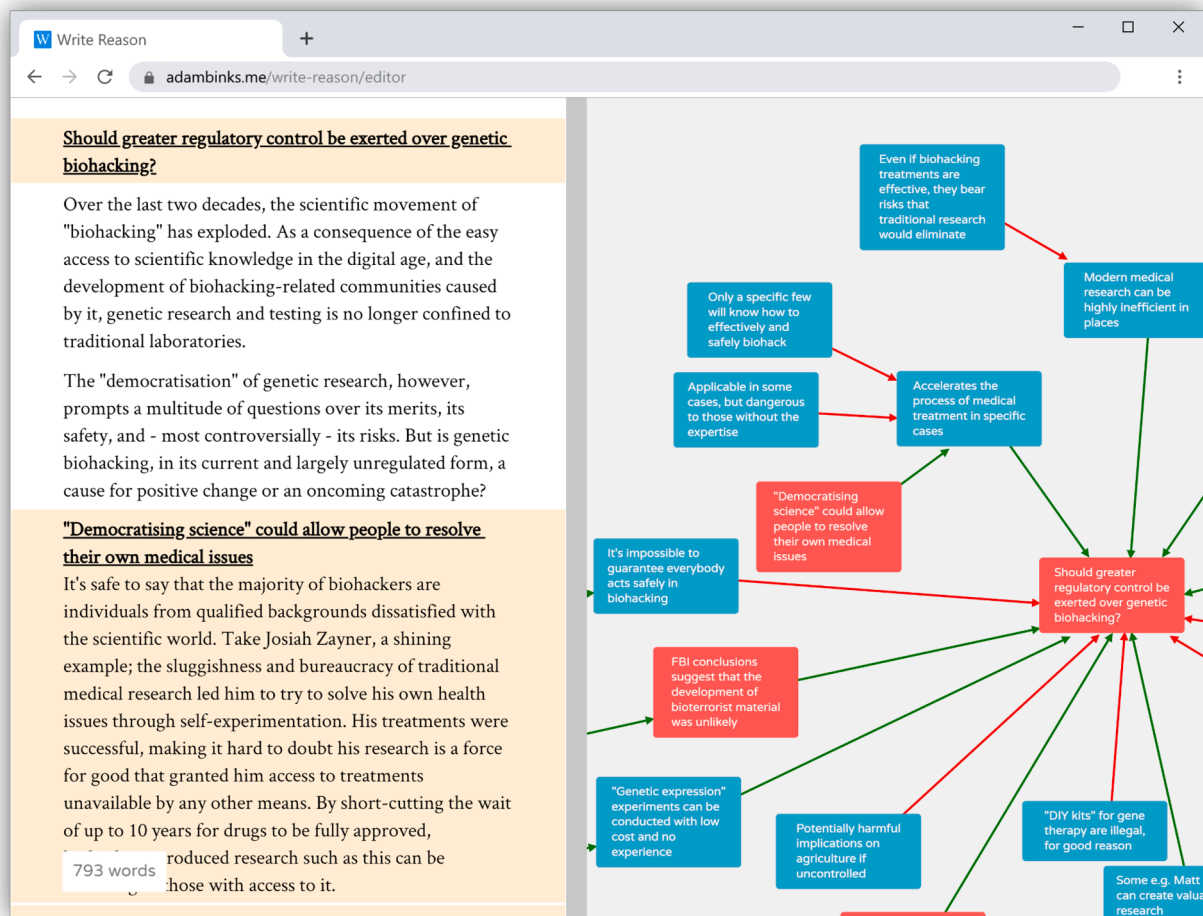


Fig. 1. An example of an essay and an argument map, built by participant 12 in our tool Write Reason. Green arrows mean 'Supports', red arrows mean 'Opposes'. Red nodes in the map are bidirectionally linked to yellow sections in the text.

the quality of their persuasive essays.

Neuwirth and Kaufer (1989) outline three ways to compare the usefulness of different external representations, in terms of: (1) information encoding (How computationally efficient and task-relevant are the mental representations produced by each external representation?), (2) storage (How useful is the chunking of information in each representation, and how often does information need to be retrieved?), and (3) goal retrieval (How does each representation aid retrieval of the writer's current and next goal?).

2.2.1. Diagrams

In essay-writing, the end product is text, which is a *sentential* representation, meaning that it is sequential (e.g., English is read left to right). However, some of the most promising representations to support the essay-writing process are *diagrammatic*: they are indexed by spatial location on a plane (Larkin and Simon, 1987). Larkin and Simon (1987) identified key advantages of diagrams: diagrams can make searching for information more efficient by spatially clustering related content (cf. Eklundh, 1992). Additionally, diagrams can make structures that are implicit in text spatially explicit, aiding their recognition (Larkin and Simon, 1987), and graphically constraining the further inferences they might support (Scaife and Rogers, 1996). For example, the argumentative support structures implicit in text can be made spatially explicit in an argument map by representing it with arrows.

2.2.2. Maps

Two kinds of diagrammatic representations have been found to

support good essay-writing: concept maps and argument maps.

Concept maps Novak's concept maps (Novak and Cañas, 2006) consist of nodes and labelled lines connecting them. Nodes represent concepts, which Novak defines as 'a perceived regularity in events or objects, or records of events or objects, designated by a label' (Novak and Cañas, 2006). Lines connecting nodes describe relationships between them. Concept maps are hierarchical: the most general concept, which is typically the subject of investigation, is placed at the top. Connecting lines are arranged to create a tree, descending from the top concept. Cross-links are the exception to the tree structure: long, labelled arrows between nodes in different areas of the map, showing relationships across domains.

Two key benefits of concept maps apply to essay-writing, building on the general advantages of diagrams identified by Larkin and Simon (Larkin and Simon, 1987). First, the hierarchical structure and short, easy-to-parse text labels make searching for information efficient (Novak and Cañas, 2006; Plotnick, 1997). Secondly, building a concept map forces the author to identify relations between concepts (Eppler, 2006), helping them discover latent conceptual structures. Finding cross-links is particularly useful, because making insightful connections between concepts in different domains is important for good essay-writing (McGinley, 1992). These benefits are supported empirically: studies have repeatedly found that students using concept maps when planning produce higher-graded essays than control groups (Azlinda et al., 2008; Jafari and Zarei, 2015; Slotte and Lonka, 2001; Zarina and Fatima, 2015). Even without training, people spontaneously produce diagrams similar to concept maps: for example, Walny et al. (2011)

found that researchers frequently used node-link diagrams when sketching ideas on whiteboards.

Though Novak originally designed concept mapping for pen and paper, software tools that allow one to move the nodes freely simplify the process, reducing the need to redraft maps as they grow. Some concept mapping-specific tools are available, such as CmapTools (Cañas et al., 2004) and R-CMap (Bar and Mentch, 2017), and concept mapping is also widely supported in general diagramming tools, including Dia (<http://dia-installer.de>), Edraw (<http://edrawsoft.com>), Microsoft Visio (<http://products.office.com/en/visio>), Visual Understanding Environment (Kumar and Saigal, 2005) and yEd (<http://yworks.com/products/yed>). Some tools aim to automatically create concept maps based on essay text, such as Concept Map Miner (Villalon and Calvo, 2011).

Argument maps While concept maps represent concepts and any type of associative relations between them, argument maps represent propositions and the inferential relations (e.g., support, opposition) between them (Davies, 2011; Reed et al., 2007). Nodes represent arguments or, in systems like OVA+ (Janier et al., 2014), argumentative concepts such as premises and conclusions. Arrows between nodes represent supporting or opposing relations between them.

Argument maps are particularly useful for planning persuasive essays (Nussbaum and Schraw, 2007). Their advantages are similar to concept maps, but where concept maps help generically identify relations between concepts, argument maps excel specifically at making the reasoning structure explicit and easy to perceive (Nesbit et al., 2019; van Gelder, 2007). This explicit representation helps users discover new arguments and responses: seeing the arguments arranged visually provides feedback to the writer about what is there and what is missing, an example of representational talkback (Schön, 1992; Yamamoto et al., 1998). Studies have found that students taught to use argument maps generate more refutations of counterarguments (Nussbaum and Schraw, 2007), and improve their essay-writing abilities more than students using traditional methods (Harrell and Wetzel, 2013; 2015) or concept maps (Fan et al., 2019; Liu et al., 2017). Lynch et al. (2014) found that a machine learning model could predict students' essay grades based on the argument maps they constructed to plan their essays. Interestingly, a model trained on automatic structural features was competitive with a model trained on expert human grading of maps, despite semantic features being ignored in the structural model. This indicates that the structure of argument maps closely relates to essay outcomes. Beside supporting essay-writing, argument maps have also been found to support critical thinking (Butchart et al., 2009; Twardy, 2004), debate analysis (Carneiro et al., 2019), argument understanding (Chiang et al., 2016; Cullen et al., 2018) and argument recall (Dwyer et al., 2010). They have been applied to group contexts such as dialogue mapping (Okada, 2008) and large scale deliberation (Klein, 2012).

Various academic and commercial software tools for argument mapping have been developed for application to many domains, such as law (e.g., AGORA-net Hoffmann, 2015, Rationale van Gelder, 2007), formal argument analysis and mining (OVA+ Janier et al., 2014), education (Belvedere Suthers et al., 1995, Reason!Able van Gelder, 2002, Athena Rolf and Magnusson, 2002), writing (the Author's Argumentation Assistant Schuler and Smith, 1992), and group deliberation (REASON Introne (2009)). Some argument mapping tools (e.g., Compendium Shum et al., 2006, CISpaces Toniolo et al., 2015) have also been developed and studied in the context of sensemaking, the task of 'creating an understanding of a concept, knowledge area, situation, problem, or work task' (Zhang and Soergel, 2014) — a task which is crucial to persuasive essay-writing, as well as many of the other domains where argument mapping has been applied. Cerutti et al. (2017) provide a good overview of argument mapping tools. Different mapping tools let users select from different sets of node and arrow types following predefined ontologies (e.g., Janier et al., 2014; Rolf and Magnusson, 2002; Toulmin, 2003; van Gelder, 2007). For example, the issue-based information system (IBIS) has three types of nodes: the 'issue', the central

question to be settled, 'positions', possible answers to the question, and 'arguments' for or against positions (Kunz and Rittel, 1970). The choice of mapping ontology has been found to affect the quality and relevance of users' arguments (Barstow et al., 2017a; 2017b; Schwarz and Glassner, 2007). In tools such as Compendium, adhering to a structured ontology allows the tool to automatically transform argument maps into other formats, such as requirement specifications or data flow diagrams (Selvin et al., 2001).

Argument mapping is a particularly promising representation for persuasive essay-writing. It preserves the general benefits of diagrams and shares the relevant benefits of concept mapping, while bringing the additional advantage of clarifying argumentative structure and thereby supporting the discovery of new arguments.

2.3. Multimedia

The studies reviewed so far show that using diagrams to plan and write essays can improve essay quality by helping writers discover arguments and relations between them. In these studies, writers often created two representations: a diagram, and a sentential essay representation. Therefore, in this Section, we review the multimedia literature, examining in general how multiple representations support complex activities.

Mayer (2002) is the main proponent of the multimedia principle: learning is more effective with pictures *and* words than with words alone. A wealth of research has supported and extended this principle (Butcher, 2014). For example, Easterday et al. (2009) found that constructing a causal diagram helped students make inferences about the effects of policy decisions, and that this effect persisted in future tasks, even those where students did not construct or use diagrams. Interestingly, Grossen and Carmine (1990) found that students with disabilities who created their own diagrams learned more than those who selected from provided ones, showing that self-authored representations can be particularly impactful. In science teaching, multimedia instructional materials can improve students' understanding of difficult concepts (Kozma and Russell, 2005), lab practices (Kozma and Russell, 2005), and problem solving (Dufresne et al., 1997). A challenge for students is developing competence in understanding, connecting and translating between different kinds of representations (Keig and Rubba, 1993; Kohnle and Passante, 2017; McCracken and Newstetter, 2001). In particular, Kozma and Russell (2005) argue that students must understand how one representation may be able to express something that another cannot.

Along with the multimedia principle itself, Mayer (2002) presents seven other principles in multimedia design. These principles are based on cognitive load theory: the limitations of working memory should inform the design of instructional materials, including multimedia (Sweller, 2011). *Intrinsic* cognitive load is the demand on working memory required by the intrinsic complexity of some information — the more interaction between elements of the information, the more working memory is needed to integrate these elements (Paas and Sweller, 2014). *Extraneous* cognitive load is the demand on working memory caused by inappropriate instructional designs which unnecessarily increase the number of interacting elements learners need to process. One of Mayer (2002)'s principles in multimedia design, the split-attention effect, identifies a common case of extraneous cognitive load: when multimedia contents are separated visuospatially, learners have to integrate them in working memory, increasing cognitive load (Castro-Alonso et al., 2019).

To mitigate the split-attention effect, a focus of multimedia research is making explicit the connections between parts of the different representations. For example, work in Magazine-Style Narrative Visualisations (MSNV) focuses on clarifying the connections between the article text and accompanying visualisations. Interfaces use interactive and visual elements to make the connection between the representations explicit, such as annotation (Lai et al., 2020), animation (Kwon et al.,

2014) and highlighting (Barral et al., 2020; Kong et al., 2014; Lallé et al., 2019; Zhi et al., 2019). By mitigating the split-attention effect, MSNV tools have been found to improve reader engagement (Zhi et al., 2019) and improve comprehension (Barral et al., 2020).

In the domain of essay-writing, few tools exist which link different representations (e.g., a concept map and the essay text). Some work looks at linking interfaces for reading and mapping, to support argument analysis and note-taking. OVA+'s (Janier et al., 2014) dual-pane view lets users build an argument map from source text, a common practice in philosophy. Similarly, MindDot (Wang et al., 2019) lets users build a concept map based on a textbook. However, to our knowledge, only a few tools bidirectionally link interfaces for mapping and essay-writing: the early hypertext tools Writing Environment (Smith et al., 1987) and SEPIA (Streitz et al., 1989), the commercial tool EssayWriter (<http://fasteressays.com>), and The Sandbox (Barzilai et al., 2020). For example, Writing Environment (Smith et al., 1987) has four panes: two mapping interfaces (a hierarchical tree, and a directed graph), and two text panes. It automatically generates the contents of the primary text pane by creating a paragraph for each node in the tree, in fixed depth-first ordering. An extension of this tool adds an IBIS-style argument mapping ontology to the directed graph pane (Schuler and Smith, 1992). Writing Environment, SEPIA and EssayWriter perform automatic transformations from map to text — which the developers of EssayWriter claim reduces complexity and allows its target audience, students with disabilities, to focus on the content rather than its presentation. Barzilai et al. (2021) found that amongst students using their tool (The Sandbox) those who revisited their maps incorporated more sources in their essays, and those who created more elaborate maps better integrated multiple arguments.

2.4. From outcomes to process

The research surveyed above finds that diagrammatic representations, such as argument maps, can improve essay-writing outcomes (e.g., Fan et al., 2019; Harrell and Wetzel, 2013; Jafari and Zarei, 2015; Zarina and Fatima, 2015), and that combining and linking multimedia can support comprehension and learning (Barral et al., 2020; McGinley, 1992; Zhi et al., 2019). Together, these results suggest that building tools that combine and link these diagrammatic representations with the sentential essay representation may be a promising approach to further improve essay-writing outcomes.

This existing work on essay-writing has largely focused on how diagram use affects essay outcomes (Cullen et al., 2018; Easterday et al., 2009; van den Braak et al., 2008). For example, Harrell and Wetzel (2013) found that argument mapping improved students' essay outcomes, in terms of their quality of evidence, counterarguments, and connections between premises. However, little attention has been given to how writers build and move between the diagrammatic and sentential representations — their processes of essay-writing with intermediate representations. Understanding these processes is important to inform the design of multi-representational tools and ensure that they support writers' workflows. Therefore, in this paper, we aim to contribute to our understanding of these processes.

3. Research questions

Our overarching research goal is to understand how different representations are used within the workflow of an essay-writing task. In particular, we examine how interaction with the representations and the flow of information between them supports completion of the task, from generation and organisation of ideas to the finalisation of the essay. A better understanding of the role of representations will, in turn, help build better future tools that support these processes more effectively. Based on our overall goal and existing research described in the previous section, we focus on three particular research questions in the domain of essay-writing:

- Q1. REPRESENTATIONS: What kind of representations do essay-writers choose to create?
- Q2. PROCESS: How do essay-writers use representations to compose an essay?
- Q3. OUTCOMES: By which mechanisms do these processes and representations support essay-writing?

4. Write Reason

To address our research questions, we designed and built a tool which combines an interface for building map representations (e.g., argument maps) with an interface to build text representations (e.g., essays). The tool, which we named *Write Reason*, allowed us to track participants' processes during essay-writing.

4.1. Design principles

Two main goals motivated the design of Write Reason: (G1) supporting integration between the textual and diagrammatic aspects of essay-writing; and (G2) capturing writers' natural essay-writing behaviours and processes. To achieve those goals, we identified three main design principles:

- DP1: *Flexibility*. Support a broad range of mapping and diagramming approaches. This would allow us to observe participants' existing strategies and avoid constraining them to a particular mapping approach.
- DP2: *Connected multiple representations*. Support easy connection and maintenance between elements in the map and in the text. Although we recognise that this adds a small element of artificiality (no such aids are generally present on current commercial mapping and word processing tools), it also greatly supports our ability to analyse how representations relate to each other.
- DP3: *Minimise cognitive load*. Minimise the need to recall information and keep the interaction as simple as possible. This is generally good practice in the design of usable interfaces but, in our case, it is particularly important so that participants are not distracted by the interface itself and focus on the cognitively demanding aspects of composing essays.

4.2. Interface design

The tool has two panes: on the left, a text document editor, and on the right, a canvas for the map (see the snapshot in Fig. 2). Both representations are displayed simultaneously to reduce demands on working memory (DP3).

4.2.1. Map

The map pane is an infinite canvas that writers can populate with nodes and arrows. Writers add nodes (e.g., Fig. 2C) by double clicking, and can directly type their contents, as well as move the nodes around the canvas through standard drag-and-drop. The canvas allows panning and zooming for navigation. The tool does not impose an ontology of node types. Writers can add an arrow between two nodes by dragging from the edge of one node to another (e.g., Fig. 2E). A modal dialog allows the writer to select an arrow type from a pre-populated list of "Supports", "Opposes" and "Expands", but it is also possible to add custom arrow types. We chose the arrow types "Supports" and "Opposes" to support argument mapping, and "Expands" gives an example of a non-argumentative relation. Along with custom arrow types, this supports the creation of most types of diagrams (DP1) and makes relations between elements in the map explicit (DP2). The interactions described above are deliberately chosen to be similar or identical to existing tools, speeding up learning and avoiding complex action sequences (DP3).

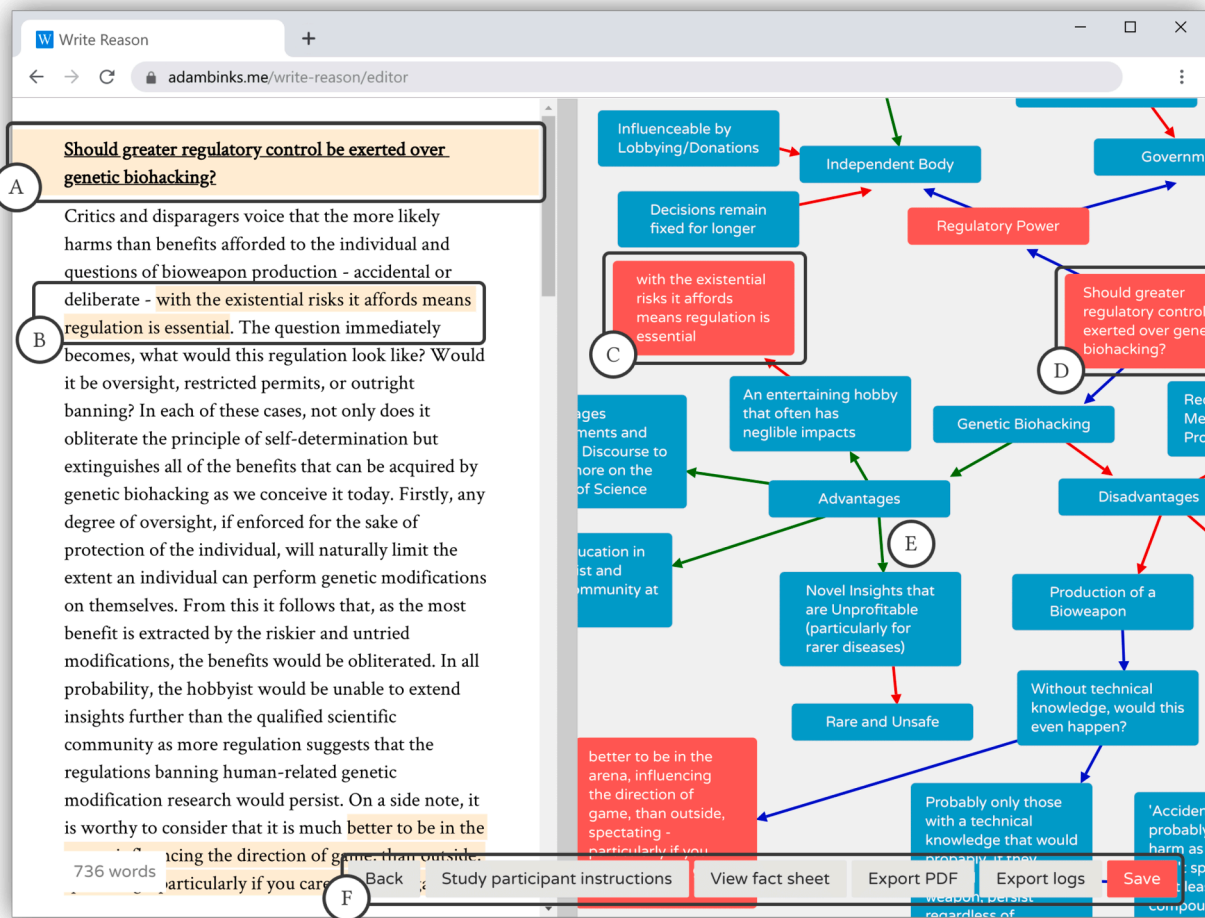


Fig. 2. An annotated screenshot of Write Reason. There are two panes: on the left, the document pane, and on the right, the map pane. Writers can connect nodes in the map (C, D) to sections of the text, either as a block (A) or inline (B). In the map pane, writers can add arrows between nodes (E). Buttons along the bottom of the screen show instructions for study participants, and enable saving, exporting and navigation (F). The screenshot shows part of participant 26's submitted work, in our study.

4.2.2. Text

The document pane contains a typical text editor. Writers can type text, as well as use the standard cut, copy and paste commands. Alternatively, writers can drag paragraphs vertically to change their order within the document. This parallels the interaction pattern to arrange nodes in the map pane and mitigates the working memory requirement of using copy and paste to reorder (DP3).

4.2.3. Map-text connections

Writers can highlight a snippet of text in the document, and drag it onto the map pane to create a connected node (e.g., Fig. 2D), where the node's content matches the highlighted text in the document. Similarly, writers can drag a map node into the document to create a yellow highlighted section (e.g., Fig. 2A). These features allow writers to both bring ideas from the map to the text, and from the text to the map (DP1). The heading of a connected section synchronises with the content of the connected node, keeping both representations up to date. Additionally, when the writer hovers over a connected node, the connected section of text lights up, and vice versa. This helps the writer keep track of the correspondence between parts of the map and text, creating connected multiple representations (DP2). The connected representations approach is reminiscent of 'brushing and linking' in information visualisation, where selected data in one scatterplot is highlighted in other scatterplots Buja et al. (1991). The writer can select the style of each section in the text: "block" style, with a heading and/or body (Fig. 2A),

or "inline" style, which is rendered inside a paragraph (Fig. 2B). When a map node is connected to a section, it is coloured red instead of the default blue, to visualise which parts of the map have not yet been added to the essay (DP3).

4.2.4. Implementation

We implemented Write Reason as a web application to facilitate access in many contexts, across different operating systems, and to facilitate remote studies. Write Reason is built with React, using Slate.js (<http://slatejs.org>) for the document pane and SVG.js (<http://svgjs.com>) for the map pane. It runs entirely on the client-side, and files are stored locally in the browser. Detailed logging of interactions allows us to analyse participants' processes. The tool records logs locally, so the data is only accessible to us when the participant explicitly shares the log file. Write Reason is online at <http://adambinks.me/write-reason/editor>, with a video overview of the interface. The source code is available at <http://doi.org/10.5281/zenodo.4781934>.

4.3. Preliminary study

As Write Reason is a new tool, before running our main study we ran a preliminary study to: a) verify that the tool is usable and fit for our purpose; b) test study materials for the subsequent main study (e.g., in terms of difficulty); c) assess the time needed to observe meaningful effects in the creation of an essay, and; d) gather an initial impression of

how people use text and diagrams in conjunction. We keep the description of this preliminary study short because it was mostly used to inform the main study described later in Section 5. We received ethical approval from the local Research Ethics Board for both studies.

4.3.1. Method

We used a within-subjects experimental design, where 24 university students (11 female, 11 male, 2 other) completed two tasks. Each task consisted of writing a short persuasive essay about one of two pre-selected questions with one of two tools. Tool is the main condition: for one essay participants used Write Reason (both map and text pane) and for the other a plain text editor (a version of Write Reason with only the text pane). The two topic questions were: *SharedSpace*: “Should shared spaces in urban planning be promoted?” and *Biohacking*: “Should greater regulatory control be exerted over genetic biohacking?”.

Before each task, we showed participants an instructional video for the corresponding tool, and gave them a chance to familiarise themselves with the tool. They then spent five minutes reading a fact sheet. The fact sheet contained the essay question and 13 snippets from popular, policy and academic articles relevant to the question. The snippets were one paragraph in length, with different content: two definitions, one example, five pieces of evidence in support of the topic, and five pieces of evidence against. To avoid giving participants a pre-made coherent argument, the fact sheet’s contents were listed in a semi-randomised order, where the lists of pros and cons were separately shuffled then combined into one list alternating between the pros and cons, with the examples and definitions inserted at fixed intervals. After reading the fact sheet participants had 15 min to write the essay, and were able to refer back to the fact sheet. We conducted the study in a controlled in-person laboratory environment: participants sat at a desk and used a vertical monitor, mouse and keyboard. We balanced the ordering of the questions (*SharedSpace*, *Biohacking*) and tools (Text Editor, Write Reason) in a 2×2 factorial design. After both tasks, participants completed an evaluation questionnaire about their experience writing essays using the tools, and comparison to other tools they had used. The materials for the study are available as supplementary materials. The total study duration was around 60 min per participant. Participants were compensated for their time with £5 Amazon voucher.

The main quantitative measurements from the study were scores of the essays. To obtain the scores we anonymised and shuffled the essay scripts and gave them to three markers trained in argument evaluation and essay marking (all three Philosophy PhD students). The markers were compensated for their time and expertise with a £70 Amazon voucher. The markers did not know which tool was used to construct any given essay. The markers scored the essays on a 1–10 scale for each of five criteria, which are described in Table 1, based on our marking guidelines. We briefly outlined these marking criteria to participants at the start of the study.

4.3.2. Results

The results of the preliminary study are shown in Table 2. On four of the five measures, we found no significant difference between essays written using Write Reason and the plain text editor. One marking criteria, persuasiveness, was significantly higher for plain text editor essays than Write Reason essays. These results indicate that, within the

Table 1
Definitions of our marking criteria.

Criterion	Description
Clarity	How articulate and clear is the presentation of claims?
Persuasiveness	How compelling is the argument?
Structure	How effectively organised is the essay?
Objection responsiveness	How well are opposing views considered and responded to?
Overall	Holistic evaluation of the essay’s quality

Table 2

Scores attained in the two conditions. WR means Write Reason, Ed means plain text editor. *t*-statistic and *p*-value are the results of dependent *t*-tests.

Criterion	Avg. WR	SD WR	Avg. Ed	SD Ed	<i>t</i> (23)	<i>p</i>
Clarity	5.79	1.55	6.1	1.34	0.98	0.34
Persuasiveness	4.4	1.67	5.31	1.57	2.34	0.03
Structure	5.46	1.53	5.76	1.43	0.93	0.36
Objection responsiveness	3.78	1.29	4.29	1.73	1.23	0.23
Overall	4.67	1.59	5.26	1.44	1.61	0.12

given constraints, students were able to use Write Reason to write essays of comparable general quality, though less persuasive, than when using a conventional editor.

The results of the preliminary study showed that the topic questions, fact sheets, and instructional materials were appropriate to prompt participants to plan and write persuasive essays. Participants were free to use the tool however they wished. When using Write Reason, 18/24 participants built a map, and 6/24 only used the text editor pane. We observed no priming effect from tool order: 3/6 participants who used only the text pane used Write Reason first, and 9/18 participants who used the map and text pane used Write Reason first. However, we noticed that the task duration (15 min per essay) was too short to generate conclusions generalisable to the academic writing context, where essays are often written over extended periods. We found that participants using Write Reason’s mapping interface often ran out of time to write the essay because they spent too much time building their map. One indicator of this was that the mean word count for Write Reason essays (221.6) was lower than for plain text editor essays (274.9). This effect is significant ($t(23) = 3.50, p = 0.002$). Participant responses from the questionnaire support this: “*the map would be useful for an actual academic essay but took up lots of time in this fast setting*”, “*I spent more time drafting my ideas on the map than I should have*”.

Having verified the study materials with this pilot study, we used these findings to design our main study. Our research goal is to understand how different representations are used within the workflow of an essay-writing task. The main study presented in the next section is exploratory in nature and therefore focuses on Write Reason only, rather than comparing it to a text-only condition.

5. Main study design

This Section presents the design of the main study, which enables us to address the research questions from Section 3. Explorations of essay-writing processes with multiple representations are relatively new. Therefore, we opted for a methodology that enabled some quantitative analysis, but that also offered qualitative analysis opportunities of the rich behavioural data to identify topics or elements of interest.

5.1. Participants

20 students (11 female, 9 male) from the University of St Andrews participated in the main study. None had participated in the preliminary study. 19 were aged 18–24, one was aged 25–34. 18 were Undergraduates, one was a Masters student, and one was a PhD student. 6 other students signed up to participate but did not complete the task or submit data, so participant numbers are between P1 and P26; we did not reallocate the numbers of participants who withdrew from the study. Participants gave informed consent.

5.2. Procedure and task

After providing consent, participants were pointed to the URL of the Write Reason web app, which guided participants through the process. The tool was the same as that of the pilot study in Section 4 with minor

improvements to its robustness. Participation was individual and fully remote. The tool guided participants to complete a demographics questionnaire and to type answers to three short pre-task questions about their understanding of essays. They then watched an instructional video which demonstrated how to use the tool. The web app then presented the task instructions and gave them access to Write Reason's mapping and writing interface.

The main task was to write a ~700 word essay in academic style on a question. As in the preliminary study, participants received a fact sheet in PDF format with wide-ranging information about the question topic. They could use the fact sheet freely to support the argumentation in their essays. The questions and fact sheets are the same as in the preliminary study described above, which we found had worked well because the topics were appropriately rich, enabled deep analysis and argumentation but, simultaneously, were not overly familiar to participants therefore avoiding high variability in previous knowledge. We gave half of the participants, chosen randomly, the *SharedSpace* question, and the other half the *Biohacking* question.

We found that the preliminary study's 15 min task duration was too short to fully exercise the tool, so in the main study we asked participants to spend 100 min on the task, distributed over up to 7 days. Write Reason allowed participants to save their progress and therefore it was possible to work on the task across multiple sessions, replicating the flexibility of university assignments. Participants kept a diary where they recorded the duration and perceived effectiveness of each session, and any usage of tools other than Write Reason. Finally, we conducted a 30 min semi-structured interview with each participant, over video call. We compensated each participant for their time with a £20 Amazon voucher.

5.3. Data and measurements

We collected the following data:

- *Task artefacts*. The essays and maps constructed in Write Reason.
- *Logs*. Detailed interaction logs captured by the tool, including each user interface action, and a version history of the essay and map.
- *Essay scores*. The primary researcher scored the essays (shuffled, before seeing the maps) on the marking criteria used in the preliminary study, described in Table 1.
- *Pre-task questions*. Paragraph-length typed answers to three questions about (1) the participant's intentions when writing essays, (2) their criteria for a good essay, and (3) what they thought were the possible purposes of writing essays.
- *Interview*. Video recording and interviewer notes from semi-structured interviews. The primary researcher asked each participant about their conception of essays, their essay-writing process, and their experience using Write Reason.
- *Diary*. A spreadsheet describing the duration and perceived effectiveness of each session spent on the task.

Analysis of the logs showed that participants spent on average 108.4 min (SD=45.6 mins) in the tool, and completed the task in 2.6 sessions (SD=1.4 sessions), with average session duration 42.2 min (SD=34.8 min).

5.4. Analysis methodologies

We used a grounded theory approach (Corbin and Strauss, 2008), supplemented with quantitative analysis, to identify key concepts in how representations are used in essay-writing. To observe participants' processes of building representations we built a player to visually step through their interaction logs inside Write Reason.

Our grounded theory approach loosely followed the Straussian method (Corbin and Strauss, 2008), moving back and forth between three stages: (1) open coding of participant maps, essays, interview

recordings and notes, pre-task questions and diaries, to identify concepts; (2) axial coding of these concepts into hierarchies, and identifying relationships; and (3) selective coding to build these relationships into theoretical frameworks. The primary researcher performed these three stages, and at every stage all of us discussed, evaluated and made changes to the codes. Over the course of many iterations, transformations emerged as the core grounded concept. We undertook a focused conceptual development (Furniss et al., 2011) to identify the properties of transformations (Q2), and how these affected essay outcomes (Q3). We also coded the functional roles played by elements of the participants' map representations and the kinds of text representations they used (Q1). Quotes in the results below are from participant interviews.

We supplemented our qualitative analysis with some quantitative analysis. We counted the occurrences of transformation and representation approaches. We also calculated the correlation between transformation and representation approaches and essay scores. Due to the early-stage nature of the study, and its non-experimental exploratory design, we did not expect this correlation analysis to find significant effects, but we included it for completeness and to suggest hypotheses for follow-up studies. We also performed quantitative analyses to identify the proportion of map content present in the text, the proportion of text content present in the map, and the order that participants moved elements from their map to their text. These methods are described in context below.

6. Results

We present the findings of our grounded theory analysis supplemented by the quantitative results. The results are organised according to our three research questions: (Q1) REPRESENTATIONS: we describe the functions played by map representations, and the kinds of text representations we observed; (Q2) PROCESS: we identified two translation processes that participants used to move between these different representations, and characterise important properties of translations; (Q3) OUTCOMES: we explored how these translations and representations supported essay-writing. The representations created by participants can be viewed interactively at <http://adambinks.me/write-reason/explore>.

When we refer to *representations*, we mean "structures in the environment that allow the learner [writer] to interact with some content domain" (de Vries, 2012). In our study, the typical structures are 2D diagrams and text created with Write Reason, although we also considered annotations on paper. In order to distinguish representations from each other, we say that *A* and *B* are different representations if they are in different media (text or map), or if they are in the same media but are visually disjoint and can be interpreted independently of other representations (i.e., they are self-sufficient). For example, P5 made two disjoint map structures in the canvas (no arrows connecting them) where both represented roughly the same set of ideas but with a different structure. One was a map of the evidence, and the other organised the same ideas to plan the order of their essay. Since P5's maps are disjoint and self-sufficient we consider each a separate representation.

Note that this is a working definition of representation that helps us describe the observed phenomena; although we did not observe cases where it was difficult to determine if something was indeed a separate representation, making distinctions might be harder in other contexts (e.g., more flexible tools that use ink, and other tasks, such as software diagramming). The fact sheet provided during the experiment is also a representation, just not one created or manipulated by the participants.

6.1. Q1) REPRESENTATIONS: What kind of representations do essay-writers choose to create?

Our first group of results concern the content and function of the

representations used in the essay-writing task. We annotated representations that appeared throughout the essay-writing process by inspecting the logs and the task artefacts.

Our initial UI design assumed that participants would create two representations (in addition to the provided fact sheet): a map and an essay text. Based on the definition and criterion above, 9/20 participants created these two representations. Contrary to our assumption, just as many participants (9/20) built three or more representations. For example, P5 created two maps and a text outline at the bottom of the editor which they consulted when writing their final essay at the top (i. e., a total of 4 representations). Finally, 2/20 participants built only one representation (an essay, but no map or outline). We first describe the kinds of map representations that participants constructed and then the text representations.

6.1.1. Map representations

Map representations varied considerably. We differentiate map representations based on how participants used the basic graphical elements (boxes, arrows, color, etc.) to represent three things: (1) ideas, (2) relations between ideas, and (3) the essay-writing process itself. Table 3 describes further distinctions in each of the three categories and summarises their prevalence across the participant sample.

Due to the importance of making connections for producing high-quality essays ('relational' essays, in the SOLO taxonomy Biggs, 1988), we now describe in greater detail how participants represented relations between ideas using their maps.

Argumentative relations The design of Write Reason was centred around mapping argumentative relations (such as support and oppose), and as Table 3 shows, these were the most common kind of relations mapped. Interestingly, we found that participants used different kinds of map elements to play this functional role: arrow colour, connection to pro/con nodes, and clustering. Examples of each of these elements are shown in Fig. 3. Some participants used arrow colour to either represent *global* argumentative relations: if a node has a green incoming arrow, it supports the essay question's central issue (e.g., biohacking regulation). Other participants used arrows to represent *local* argumentative relations: if the arrow from A to B is green, A supports B, or if it is red, A opposes B. Representing local argumentative relations has the advantage that participants can record local counterarguments, even if they are not directly a pro or con of the global essay question. Of the 13 participants who used arrow colour to represent argumentative relations, 4 took the *global* approach, while 8 took the more expressive *local* approach. P23, who represented local argumentative relations, reported that "I used red arrows when a point directly contradicted another."

Table 3

Overview of the functions for which participants used elements of their maps, and the number of participant map representations which did so.

Function	Description	Maps
<i>Representing ideas</i>		
Fact sheet content	Nodes to represent ideas from the fact sheet.	17
Original ideas	Nodes to represent ideas not on the fact sheet.	5
Issues	Nodes to represent questions to be settled (termed an 'issue' in IBIS Kunz and Rittel (1970)).	14
<i>Representing relations</i>		
Argumentative relations	Arrows, arrow colours or clustering to represent argumentative relations between ideas.	16
Provenance	Nodes or arrows to represent the source of their ideas.	6
Other relations	Arrows to represent other semantic relations (e.g., 'is an instance of').	12
<i>Representing the process</i>		
Planning essay order	Nodes or arrows to represent the planned order of mapped ideas in their essays.	3
Representing task reqs	Nodes to represent task requirements (e.g., target word count and time limit).	1

"This was really useful because it helped me see where things directly responded to each other."

Provenance relations Participants used various elements of their maps to represent the source of their ideas: some used special nodes and arrow colours (e.g., Fig. 3B), while others listed sources in node labels, like an in-text citation. These elements acted as an *index* to the fact sheet: reminders of where to find pieces of information as described by P7: "The graph worked as a shortcut to know where I was going back to in the info sheet. [...] I knew where to look for the relevant information on the info sheet, so I could just go straight there and copy a couple of things."

Other relations 12/20 participants used arrows to represent semantic relations between connected nodes other than argumentative and provenance relations. For example, P3 connected their nodes 'Functions' and 'Reduce vehicle speed' with one of the predefined blue 'Expands' arrows, to show that reducing vehicle speed is one of the functions of shared spaces (Fig. 3C).

6.1.2. Text representations

Our analysis identified three kinds of text representations: essay, fact sheet reproduction, and essay structure list. See Table 4 for definitions and counts of each. We observed the greatest variety of media in fact sheet reproductions. Four participants pasted part or all of the fact sheet into Write Reason's document pane, to easily see it and their map side-by-side. In their interviews, others reported using different media for similar representations: P14 printed the fact sheet and highlighted snippets on paper, P21 wrote paper notes of relevant fact sheet snippets, and P22 copied relevant snippets into a Word document. As Table 4 shows, only one participant, P5, planned their essay in text using bullet points, while some other participants used their map to play a similar role (e.g. Fig. 3D).

6.1.3. Summary: Representations

We identified three high-level functions for which participants used elements of their maps and three kinds of text representations. Most participants built maps. Of these, most used nodes to represent ideas from the fact sheet and other 'meta' nodes to structure their map. Most participants connected nodes with arrows to indicate argumentative relations, or in some cases, non-argumentative relatedness and provenance. Some participants also represented whether nodes supported or contradicted the central issue by clustering nodes, and others by connecting them to 'pro' and 'con' nodes. Notable minority strategies included mapping out original ideas, using the map as an index or shortcut to relevant fact sheet snippets and using the map to plan the order of the essay. Regarding text representations, seven participants copied part or all of the fact sheet as a separate textual representation and one created a text outline.

6.2. Q2) PROCESS: How do essay-writers use representations to compose an essay?

Having described the various map and text representations participants built, we now turn to our central question of how participants created these representations and how information moved between them throughout the essay-writing process.

A key concept that we needed to operationalise for our analysis is *transformation*. We identify a transformation from representation A to representation B when we observe that A informs the content or structure of B. In our context, common examples of transformations include using information from a map to create the essay text, or using information from the fact sheet to create a map. As previously introduced in Section 6.1, we observed that almost half of the participants (9/20) built more than two representations (in addition to using the provided fact sheet for a total of 4+ representations), and therefore performed three or more transformations in the course of the task. In total, we observed 49 transformations.

Transformations can either be in-place, where the new representa-

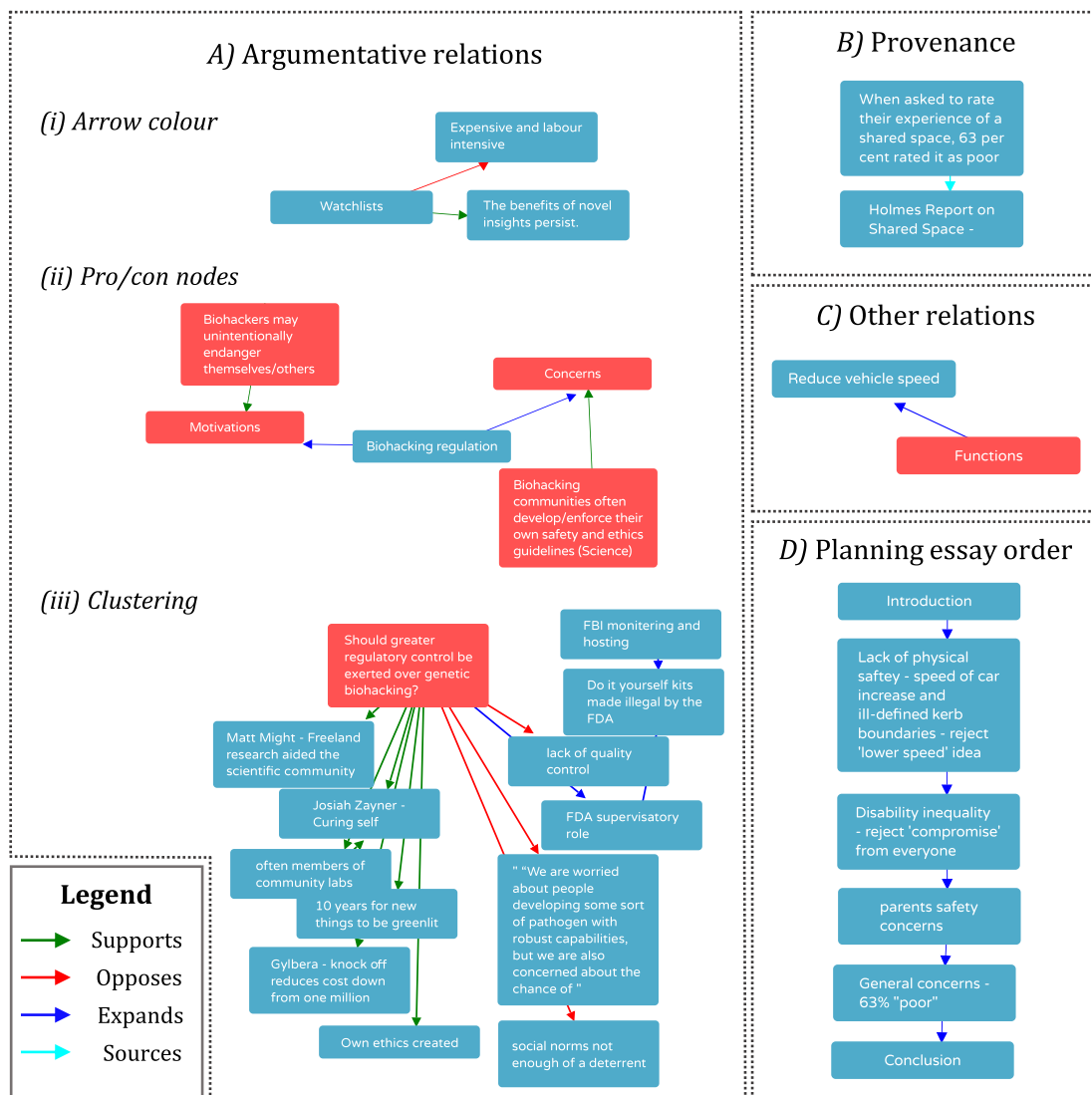


Fig. 3. Examples of elements participants used to represent (A) argumentative relations, (B) provenance, (C) other relations and (D) planned essay order.

Table 4
 The types of text representations we observed, and the number of participants who built each type of representation.

Text representation	Description	Count
Essay	A textual essay: the expected end result of the task.	20
Fact sheet reproduction	A reproduction of some or all of the provided fact sheet.	7
Essay structure list	A bullet-style sequence of summaries of planned paragraphs	1

tion *B* replaces or overwrites the old representation *A*, or not in-place, where *A* and *B* exist separately, side-by-side. All 49 transformations we observed were not in-place. We refer to not in-place transformations as *translations*. A representation can be informed by multiple other representations; hence, a representation can be involved in multiple translations.

During the design of Write Reason, we implicitly assumed that participants would move from map to text by creating informationally equivalent representations (Simon, 1978). This is a special case of translation, which we term *transliteration*. We expected writers to intend to maintain consistency between their map and text. This would have been consistent with Larkin and Simon's observation that, when

presented with a sentential description of a pulley problem in physics, most people immediately translated it to an informationally equivalent diagrammatic sketch of the described scene (Larkin and Simon, 1987). By examining whether participants' map and text representations were informationally equivalent, however, we found that this was not the case.

For each of the 18 participants who built a map, we annotated which ideas in the final essay text were also represented in their final map and, conversely, which ideas in their final map they also represented in their essay text. As expected, we found that there was a core part of essay content that corresponded to map nodes: on average, participants referred to the content of 86% of their map nodes in their essay text ($SD=21\%$). Conversely, 74% ($SD=17\%$) of words in the essay text were directly related to the content of any map nodes. It follows that 26% of words in the essay text were not directly related to the content of any map nodes and 14% of nodes were not directly related to the essay text. All 18 participants' essays contained some information that was not present on their maps; i.e., maps and text were not informationally equivalent. Therefore, none of the participants who built a map and text performed a transliteration between them. Instead, they used translations that changed the information represented.

Before proceeding with further analysis of how translations characterise the essay-writing process, we will first note a subtlety of infor-

mation exchange in translations. As defined above, a translation occurs when elements of representation *A* inform elements of representation *B*. We observed that information also mainly flowed from *A* to *B*. For example, a node in map *A* might inspire a sentence in text *B*. However, we found that information sometimes also flowed in the opposite direction, whereby elements of the old representation *A* were modified during the creation of the new representation *B*. These changes were often very minor. For example, participants made small adjustments to their maps while writing their essays, such as correcting typos, adjusting quotes or importing one or two new nodes. These cases show that a translation can also change the starting representation *A*.

6.2.1. Properties of translations

Our analysis revealed three observable properties of translations, which are summarised in Table 5 and presented in more detail in the rest of this Section.

Change in representation type

This property captures the move between one kind of representation to another (as introduced in Section 2.2.1). Most commonly, we observed 19 translations from sentential to diagrammatic representations: typically, the move from the fact sheet to a map. We also observed 18 translations from diagrammatic to sentential: typically, the move from a map to the essay. For example, P10 first used the fact sheet to build a map, then used this map to write their essay. More unusual cases include 3 diagrammatic to diagrammatic translations, such as P11’s move from a map of the evidence to a map planning their essay’s structure, and 9 sentential to sentential translations, such as P5’s move from their bullet-style text structure list to their essay.

Cardinality

When information moves from one representation to another, the translation process might involve different levels of granularity. For example, we often observed a single item of information in a representation being transformed into several items in the destination representation. We call this the cardinality of the translation, of which there are three types:

- *One-to-many*: One element in the old representation corresponds to multiple elements in the new representation. We observed 21 translations in which selected elements of the original representation generally turned into many elements in the destination. For example,

Table 5

Overview of observed properties of translations. Numbers in parentheses indicate (1) the number of participants’ translations with that value, out of the total 49 translations, (2) the number of participants who performed translations with that value. Note that six translations were to or from representations built outside Write Reason (two on paper, one MS Word document), which we did not collect, so cannot determine some of their properties.

Property	Description	Possible values (translations, participants)
Change in representation type	Change in kind of representation from old to new.	Diagrammatic to diagrammatic (3, 3) Sentential to diagrammatic (19, 17) Diagrammatic to sentential (18, 18) Sentential to sentential (9, 9)
Cardinality	Correspondence between elements of the old and new representations.	One-to-many (21, 13) One-to-one (22, 15) Many-to-one (0, 0)
Explicitness	Change in comprehensibility to a general audience.	More explicit (12, 11) Less explicit (15, 14) Same explicitness (16, 12)

P11 moved from map to essay in a one-to-many translation. A single node in their map (e.g., the node ‘parents safety concerns’), corresponds to a full paragraph of sentences in their essay (e.g., a paragraph containing a quote from a concerned parent, and an argument about the danger of low kerbs and children’s lack of understanding).

- *One-to-one*: One element in the old representation corresponds to at most one element in the new representation. Note that this does not require every element in one representation to correspond to an element in the other (i.e., some elements may not be translated). We observed 22 one-to-one translations. For example, P10’s map node ‘FDA has a role in public education and engagement for potential health risks (Science)’ corresponds to one sentence in their essay: ‘Likewise, the FDA is also involved in public education about and engagement with potential health risks’.
- *Many-to-one*: Multiple elements in the old representation correspond to one element in the new representation. We did not directly observe many-to-one translations in the task data, but reports in the semi-structured interview indicate that participants use these translations in a typical essay-writing process when a fact sheet is not provided. P17 described their typical process: first, they read sources, making detailed notes on them in a MS OneNote text file. The next step is a many-to-one translation to a MS Word document: ‘I spend time condensing that [MS OneNote text file] as much as I can — I like to get it all on one bit of paper’.

Explicitness

In this context, the explicitness of a representation concerns how comprehensible it is to a general audience. A representation sits somewhere on the spectrum between what we call an *explication* and a *pointer*. An *explication* is comprehensible to a public audience: it fully represents its own meaning, so anyone able to read the language can interpret the intended meaning. For example, a good introduction in an essay might contain an explication of the topic, where uncommon terms are carefully defined using non-technical language. At the other end of the spectrum, *pointers* are for a private (individual) audience. A pointer uses a very concise representation, intended to remind its author of a fuller thought in memory — and is probably incomprehensible to someone who does not have the same association in their own memory. For example, P26 made a map node labelled ‘Existential risk weighting?’, which is difficult to interpret for a general audience, but aimed to serve as a useful reminder for P26.

With these definitions in mind, we identified that a property of translations is the shift along this pointer-explication spectrum. For example, P11’s translation from their map to their essay shifts towards explication, moving from pointers like ‘parents safety concerns’ to full paragraph explications. Some translations shift in the opposite direction away from explication, such as P3’s transformation from the fully explicated fact sheet, to a map where nodes were pointers like ‘Save 50%, delays 66%’. Other translations involve little or no change in explicitness. For example, in P8’s map, nodes are labelled with full paragraph explications of the ideas, so P8’s translation from their map to essay involved little change in explicitness. In total, we identified 12 translations involving a shift to more explicitness, 16 maintaining roughly the same level of explicitness, and 15 shifting to less explicitness.

6.2.2. How representational translations unfold

As highlighted above, we view the general process of essay-writing as the creation and evolution of representations that often relate to each other through different types of representational translations. This process starts with sources (in this case, the fact sheet), and ends with an essay. To go from sources to essay, writers use a chain of one or more translations. Some participants (2/20) made only one translation, going directly from the fact sheet to the essay. More commonly, 9/20 participants made two translations, from the fact sheet to a map, then from map to essay. Other participants made three (7/20) or four (2/20)

translations, such as transforming from a first draft map to another map, or to a bullet point sentential representation, before transforming this to their essay.

Translation is not the only process at play, since representations also often grow and change internally. We observed two main ways in which these two processes of internally changing representations and translating between them unfolded in time:

- **Batch translation:** The participant built an entire representation (e.g., a map), then translated it into another representation (e.g., an essay).
- **Interleaved translation:** The participant built multiple representations in parallel, while often switching to translating information between the representations.

Batching was by far the most popular process, used by 17 participants in 23 translations, while interleaving was rare, used by only 5 participants in 5 translations. Three participants used interleaved translations to continue extending the map while writing the essay text (e.g., Fig. 4). Two other participants interleaved the creation of a small fact sheet reproduction (a single paragraph copy-pasted into the text pane) during the translation from a map or fact sheet to another map. Note that translations from the fact sheet are neither batch nor interleaved transformations, as participants did not create the fact sheet. Participants used batching for all other translations we observed.

We were also interested in the order in which participants moved ideas from their maps to their essay texts. This is a specific type of translation which provides insight on how a diagrammatic structure (the

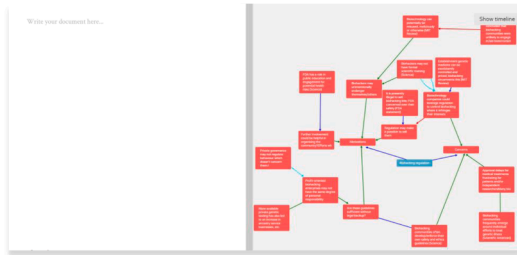
map) is linearised into a sentential representation. Diagrammatic representations in Write Reason are graphs, and traversals are an established way to linearise a graph (Kleinberg and Tardos, 2006). Therefore, we compared the *actual* order participants moved ideas from their map to their essay text with the *predicted* order according to two common graph traversal algorithms: depth-first and breadth-first traversal. We aimed to understand the extent to which the ordering of translations depended upon the structure of the old representation.

For each participant, we annotated the order in which points described in nodes on their map appeared in their final essay text. We then analysed how closely this actual translation order corresponded to a depth-first or breadth-first traversal of the map. To do so, we built functions implementing the depth-first search and breadth-first search algorithms. Let $N_1, N_2, N_3, \dots, N_n$ be the actual order that nodes were transformed from map to text. Given a node N_k and the previously visited nodes $\{N_i | i < k\}$, the depth-first and breadth-first functions return the set of next possible nodes P , which is N_k 's non-visited children or siblings respectively. Note that we allow any child/sibling rather than the left-most one, as a right-to-left or shuffled traversal order is equally valid for these purposes. We then compared if N_{k+1} is in P , to see whether the actual next node matches a next node that each ordering algorithm might visit. This yielded the percentage of nodes each participant transformed in depth-first and breadth-first order in a map to text translation.

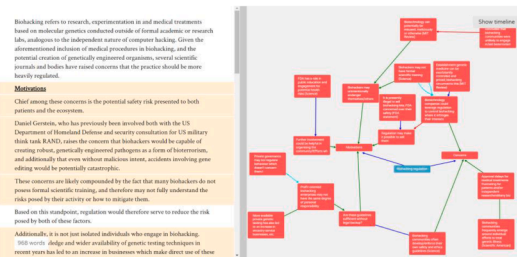
After removing participants who did not build maps (2/20), and participants who built maps not connected through arrows (2/20), we found that a majority of existing maps were translated to text in a

Batch translation

1. Create map



2. Translate map to text



Interleaved translation

Create part of the map, translate it to text, repeat

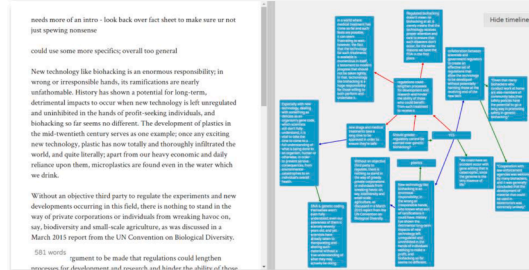
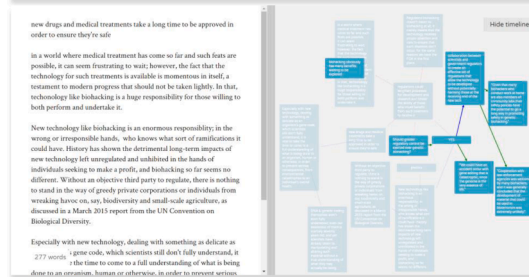
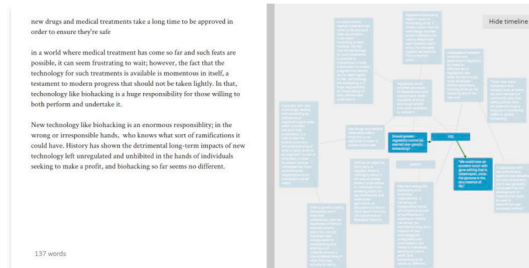


Fig. 4. Screenshots from P10's batch transformation and P8's interleaved translation. Faded nodes haven't yet been created. The timeline view is available at <https://adambinks.me/write-reason/explore>.

fashion that resembled more closely a depth-first traversal (9/16 participants) and the remaining 7 followed each equally, but also to a very small degree. On average, depth-first explains 55.3% of the linearisation process ($SD=25.7$) and breadth-first explains 39.3% ($SD=21.8$), in the 16 examined participants.

6.3. Q3) OUTCOMES: By which mechanisms do these processes and representations support essay-writing?

In the previous two Sections we have characterised the representations that participants created and the processes by which participants moved between them. Now we attempt to establish connections between the different phenomena that we observed (e.g., the use of different graphical elements, the use of interleaved vs. batched translations) and the essay quality as measured by the marked scores. The scores are summarised in Table 6.

The Pearson correlations between the 33 different coded categories and the 5 scores (a total of 165 correlations) are shown in Fig. 5. At a standard statistical significance level ($\alpha = 0.05$) there is a positive correlation between the use of nodes in the map to represent original ideas and the persuasiveness and structure scores of the essay, between the use of one-to-many translations and all scores except clarity, and negative correlations between using a map to plan essay order and the use of balancing arguments (the Objection Responsiveness score), and also a negative correlation between one-to-one translations and the structure score of the essays. Note that, as expected given the high number of tests, controlling for the false discovery rate (e.g., with Benjamini and Hochberg's procedure Benjamini and Hochberg, 1995) does not yield any statistically significant results.

7. Discussion

This section interprets the most important findings in the context of the essay-writing task and the broader literature, first by addressing the questions stated in Section 3, then more general themes. We then describe the limits on the knowledge gained and offer open questions for future research.

7.1. Q1. REPRESENTATIONS: What kind of representations do essay-writers choose to create?

Participants adopted very different approaches when creating representations. Within the maps created, there were important differences in how participants used the nodes and arrows to represent different kinds of elements. There was no single dominant style of using graphical elements, nor do we have evidence to suggest one way is better than others, especially as people might differ in cognitive style, essay-writing experience, and topic knowledge. In many cases the map served as a melting pot of ideas from the provided literature (the fact sheet), statements of the issues or questions being discussed, and original ideas that occurred to participants during the process.

A large majority of participants (16/20) used the map to represent argumentative relations between ideas, a form of argument mapping (Davies, 2011). Participants seem to have understood that simply connecting ideas to each other, without qualifying the nature of their connection, would not be sufficient to help them plan their essay. They

Table 6
Summary of participant essay scores.

Marking criterion	Mean	Median	SD
Clarity	7.00	7.0	1.86
Persuasiveness	6.60	6.5	1.98
Structure	6.40	7.0	1.57
Objection responsiveness	6.85	7.0	1.84
Overall	6.75	7.0	1.74

found different ways to represent these relations, such as using arrow colour, spatial clustering, and connection to nodes representing argumentative stances on the issue (pro and con nodes). These approaches offer different trade-offs around simplicity, ability to provide an overview, navigability of the argument structure, and even the ability to develop new ideas from existing ones.

Six participants used the map to represent the provenance of ideas in the fact sheet, effectively making their map representations an index of the fact sheet. This suggests that, at least in some cases, when participants built representations they were already aiming to minimise the need to consult other representations by keeping track of highly essay-relevant information. Participants also used nodes and arrows to represent other kinds of relationships, such as conceptualisations of ideas (e.g., Fig. 3C). Interestingly, only three participants used their maps to outline their intended essay structure. This was one of the scenarios that we had in mind when designing Write Reason, but it was neither popular nor evidently useful (see also Section 7.3 below). This might have been because the essay was not long enough to require cognitive aids for the structure, or because participants thought that the text essay itself was sufficient to grasp its general structure.

Participants also used text representations in multiple ways, although we observed less variability in sentential representations than in the diagrammatic representations discussed above. Participants used text to write the essay, to create selections of the fact sheet data and one also used the editor to create an outline. A tool that supports text structures like hierarchical outlining and bullet points might yield a greater variety of sentential representations.

While we designed Write Reason to support one map and one text representation, many participants went beyond this. Without any explicit prompts, nine participants (out of 20) created three or more representations. This also relates to the point above about the variety of representations; different representations, even if they are in the same media (e.g., diagrammatic) might support different parts of the process. We suspect that many participants recognised these potential affordances and that led them to create multiple representations in the same media.

7.2. Q2. PROCESS: How do essay-writers use representations to compose an essay?

Our analysis of how the representations were constructed and how they relate to each other offers some interesting insights. We found that information does not travel from one representation to another in a simple, one-to-one fashion. Instead, we observed that independent units in one representation became multiple units in another, such as a single idea represented in a map node being expanded into multiple elements in the text (or in another representation). This reflects a natural role of intermediate representations to serve as scaffolds and indexes to the understanding of source materials that the writer holds in their head, or to their previous knowledge. As the process unfolds across representations, some elements disappear (i.e., they do not get translated to the subsequent representation, a kind of filtering) and some explode into multiple elements. In this way, translations between representations play a key role in the selection and development of ideas.

Interestingly, we did not observe any cases of many-to-one translations. We expected to see this in the form of summaries or abstractions (e.g., an item in the new representation providing a reference for a group of more specific cases in the original one). It is possible that the nature of the task or the experimental setup did not require these kinds of operations. It is also possible that these processes are cognitively harder to support through representations and translations or that these relationships are naturally represented in the groupings and the order of the final text essay.

Studying the sequences of changes in the representations that participants created allowed us to identify two main ways to create representations: batching and interleaving. In batching, the more dominant

(a) Correlations between scores and properties of map and text representations.

Property	Clarity	Persuasiveness	Structure	Objection resp.	Overall
Used graph	0.18	0.19	0.31	0.25	0.34
Used paper	0.0	-0.19	-0.09	0.03	-0.15
Fact sheet content (quoted)	-0.11	-0.21	-0.26	-0.14	-0.15
Fact sheet content (paraphrased)	0.06	0.04	0.21	-0.07	0.12
Fact sheet content	0.0	-0.09	0.02	-0.04	0.02
Original ideas	0.06	0.48*	0.45*	0.18	0.36
Issue node	0.18	0.15	0.03	0.25	0.29
Argumentative relations	0.21	-0.04	0.13	0.31	0.22
Arg relations: arrow colour	0.12	0.34	0.12	0.29	0.39
Arg relations: pro/con connected	0.11	-0.09	0.05	-0.05	0.12
Arg relations: clustering	0.17	-0.03	-0.03	0.37	0.1
Provenance	0.0	-0.15	-0.1	0.12	0.03
Other relations	-0.06	0.09	0.15	-0.3	0.12
Planning essay order	-0.08	-0.28	-0.2	-0.51*	-0.27
Representing task requirements	0.0	-0.07	0.09	0.15	0.03
Essay structure list	0.0	-0.07	0.09	0.15	0.03
Fact sheet reproduction	0.17	-0.07	0.01	0.29	-0.08

(b) Correlations between scores and properties of translations.

Property	Clarity	Persuasiveness	Structure	Objection resp.	Overall
Number of translations	0.27	0.05	0.22	0.19	0.12
Interleaving	0.13	0.06	0.15	-0.14	0.02
Batching	0.19	0.15	0.22	0.32	0.19
Diagrammatic to diagrammatic	0.23	0.09	0.16	-0.28	0.06
Sentential to diagrammatic	0.06	0.14	0.22	-0.01	0.1
Diagrammatic to sentential	0.18	0.19	0.31	0.25	0.34
Sentential to sentential	0.11	-0.23	-0.17	0.36	-0.16
More explicit	0.09	0.04	-0.04	0.37	0.05
Same explicit	0.26	0.26	0.24	0.01	0.31
Less explicit	-0.05	0.14	0.06	-0.09	0.04
One-to-many	0.4	0.57*	0.53*	0.47*	0.52*
One-to-one	-0.25	-0.38	-0.46*	-0.39	-0.33
Proportion of map nodes in text	0.11	-0.14	0.07	0.09	0.09
Proportion of text words in map	0.01	-0.07	0.13	-0.22	0.04
Depth first proportion	-0.08	-0.15	-0.22	0.13	0.03
Breadth first proportion	0.07	-0.11	-0.39	0.29	0.12

Fig. 5. Pearson correlation coefficients between properties of participants' representations and translations and the scores. * indicates statistical significance ($p < 0.05$) before applying the false discovery rate control (Benjamini and Hochberg, 1995). Stronger positive correlations are highlighted with more saturated red backgrounds while stronger negative correlations are highlighted with more saturated blue backgrounds. No results were significant after applying the false discovery rate control.

approach, participants built one representation first and then used it to create another. In interleaving processes, participants started translating content to other representations during the construction of the original. These two styles have different affordances and are likely to result in different cognitive loads. Interleaving requires more changes of representational context; every time that one item is carried over to the destination representation, the writer has to bring back to memory its different working conventions and mappings. This might be particularly taxing when the two representations are in different media. For example, rapidly moving between an argument map and an essay would involve switching between writing for a personal and an external audience, and between fitting elements within a spatial argumentative structure and a linear text flow. We suspect that interleaving translations are thus more cognitively demanding and leave fewer mental resources for the writer to perform higher-level cognitive tasks such as summarising, abstracting or generating counterarguments, leading to a shallower multi-structural essay, in the SOLO taxonomy Biggs (1988). Our results show some limited support for this: we observed that interleaving processes generally result in more superficial argumentation in essays, and the objection responsiveness metric correlates negatively with interleaving and positively with batching (although not statistically significantly; see Section 7.3).

We were also very interested in how the non-linear structures of diagrammatic representations (the maps) were translated to more linear representations (specifically, the final text of the essay). The results suggest that this process is not strongly guided by an orderly navigation of the graph that resembles algorithmic traversal. Although the actual behaviour of participants is closer to depth-first traversal than breadth-first traversal, most of the time participants did not stick to one, or just did not follow either. The preference for depth-first is consistent with the lower memory requirements of this algorithm (in depth-first, there are fewer paths not fully travelled to keep track of Korf, 1985). Another factor could be that depth-first places connected nodes closer together in the essay, which may reduce the demands on readers' working memory.

Another key finding was that, of the 49 translations we observed, none were transliterations (translations that ended with two informationally equivalent representations). In our approach to measure

informational equivalence, it would be easy to obtain a text from a map through a simple linearisation process of the nodes in the map with additional connectives. Such a process can further be supported by a tool such as EssayWriter ([fasteressays.com](https://www.fasteressays.com)), which automatically translates between map and text. Even in this translation, the structure of diagrammatic and sentential representations differ and may not be considered a transliteration. It is generally hard to make two representations totally informationally equivalent as, for example, in Larkin and Simon's diagrams of physics problems (Larkin and Simon, 1987). One reason is that a text and a map have unavoidably different spatial relationships between their constituent subelements: in a text any given paragraph or sentence is immediately preceded by a single sentence and followed by one sentence (i.e., has a linear form), whereas in a map or diagram a box representing an item might have multiple neighbours, not just two, even if the box is not explicitly connected. Transliterations may also be undesirable — moving all information element by element in a way that only results in a different form of identical content might be perceived by writers as a waste of time.

7.3. Q3. OUTCOMES: By which mechanisms do these processes and representations support essay-writing?

The results of our analysis are least conclusive about Q3, which is also the hardest to address. There are two key methodological issues here, derived from the original intent of the study and the study design choices. First, the study was designed to explore the elements and processes that might be of importance regarding different representations in the writing of essays. The main approach is qualitative and exploratory, which means that we did not select a specific model of essay-writing or representation *a priori* and, instead, we used an open coding approach that would allow us to discover the different concepts and processes that might be of importance. Therefore, trying to infer causal relations between the observed phenomena and the essay scores is premature. Second, our analysis coded, classified and categorised a large number of elements and behaviours (33 categories), which can be correlated with the five different scores from the marking process (i.e., a total of 165 correlations). With the size of our sample and the natural variability in individual strategies that can be expected from a complex task, it is

unreasonable to expect strong statistical correlations, especially if we want to control the false discovery rate for the large number of factors. Hence the strong or moderate correlations in the analyses that we found should be interpreted as indicators that hint at interesting possible effects, rather than as direct evidence of the effect of certain participant practices on the quality of the essays. Further causal links would have to be supported through experiments that exercise at least some control on what participants do (e.g., through training).

With the caveats described above in mind, we would like to highlight two ideas that are suggested by the results summarised in Fig. 5 and our experience in the qualitative analysis and interpretation of the results. First, we believe that creating intermediate representations (other than text) to construct essays has value. Participants who chose not to use diagrammatic representations (i.e., those who directly decided to go to text: P17 and P25) did not create particularly strong essays (6 and 4 respectively in overall marks). Conversely, some of the best essays were by participants who built multiple representations (e.g., P6, P26). Although the evidence of correlation between scores and the number of transformations (and hence, the number of additional representations) is not definitive, the correlations are positive for all scores. Other research has shown benefits of diagramming of different types (e.g., concept maps Novak and Cañas, 2006, argument maps Harrell and Wetzel, 2013 and sketches Cherubini et al., 2007).

Second, we interpret that the positive correlation between scores and one-to-many translations and the negative correlation between scores and one-to-one relations indicates the value of translations that develop ideas. There seems to be value in using intermediate representations to not just represent the same items, but to process and transform the information as it moves from one representation to another.

7.4. Multiple representations: To map or not to map

Creating multiple representations adds obvious additional cognitive and time costs to the process of crafting an essay. P21, who made little use of the map, described this cost: “I didn’t need [to build a map] to write the essay, so it felt like if I did that it was doing extra things that I didn’t need to do”. More specifically, it takes effort and time to create another representation, to switch attention between one representation and another when creating or reading it and, possibly, to keep representations consistent with each other.

Yet, most of our participants seem to have realised that there are benefits in creating multiple representations that are likely to outweigh the added cost; some participants created many. Creating multiple intermediate representations is not a pragmatic action (it does not directly contribute to goal of writing the essay) and is instead an epistemic action: it aims to improve the writer’s internal environment, perhaps by reducing the space complexity, time complexity, or chance of error in future mental operations (Kirsh and Maglio, 1994). This is an application of Zhang’s notion of representational determinism (Zhang, 1997), which is well supported through evidence (e.g., Larkin and Simon, 1987; Zhang, 1997; Zhang and Norman, 1994). If the format of a representation determines the information that is perceived, the processes that are activated and the structures that can be discovered, it stands to reason that additional and more diverse representations would support essay-writing by allowing writers to access more ideas, discover alternative relationships between ideas, and to build these into more persuasive schemas. For example, separating pros and cons spatially in a map and linking ideas through lines can enable faster navigation and access to the right information (as shown before, e.g., in Blackwell, 2002; Nardi, 1993). This is consistent with several of Multimedia Learning Theory’s principles, which state that learning is affected by the modality and combinations of media that is used to express the information (Mayer, 2002; Moreno and Mayer, 1999; Sorden, 2012).

The mechanisms by which multiple representations (or media) support the task are likely richer than the individual benefits of specific representational forms. For example, we observed that P6 was able to

generate, preserve and then integrate more of their original ideas in the essay by creating multiple representations that separated the materials in the fact sheet from their own ideas. Hence, having multiple representations can, in itself, provide the benefit of separation of concerns.

7.5. The space between representations: Complementing Zhang’s representational Determinism

In this Section we bring our results about participants’ workflows (reported in Section 6.2, discussed in Section 7.2) into the larger context of existing research, highlighting the significance of what we consider our key contribution: bringing attention to transformations.

The preceding subsection (Section 7.4) and the Literature Review (Section 2) show that existing literature on representations for complex tasks (e.g., comparisons between diagrams and text, visual vs. textual programming) focuses mainly on the affordances of the representations themselves. This ranges from general findings about representations (Zhang, 1997; Zhang and Norman, 1994) and diagrams (Larkin and Simon, 1987) to evaluation of specific techniques such as argument maps and concept maps in tasks closer to our own (e.g., Fan et al., 2019; Harrell and Wetzel, 2013; Slotte and Lonka, 2001; Zarina and Fatima, 2015).

Our own experience with Write Reason, and the analysis of our study of essay-writers points to a contiguous but mostly overlooked aspect: the process of information exchange between representations (which we call transformation). In other words, much of the important cognitive processing does not happen with or within a single representation, but instead in the processes that move information *between* multiple representations.

This space between representations is particularly rich precisely because origin and destination representations can be of very different types, resulting in many combinations (e.g., ideas diagram to essay text, outline to map of essay structure, etc.) This leaves many interesting and potentially important questions open. For example, when the destination representation is text, is a diagram explored and navigated in a specific way? Are certain combinations of origin and destination representation pairs more appropriate for certain processes (e.g., which combinations better support creative convergent/divergent processes Cropley, 2006; Runco and Acar, 2012 or convergent/divergent arguments Walton, 2005). Additionally, the temporal aspects of translations (batching and interleaving) further point to radical potential differences in the processes that might have important cognitive implications. For example, interleaving, which leads writers to see the same idea in different representations and related to other ideas almost simultaneously might result in more complete exploration of the adjacent ideas. Conversely, a batching process likely allows the writer to have a better overview of the idea field, which could facilitate the process of synthesising and rearranging information.

The discussion above is important because it points to key issues of working with externalized representations. These can have implications for how to teach students to build arguments and how to build the next generation of multi-representational interfaces. Importantly, we believe that closer consideration of transformations between representations opens an opportunity for interface-driven support of more complex workflows or pipelines that are not possible with pen and paper externalizations or the current (relatively simplistic) interface support from writing tools. For example, when we asked about their typical essay-writing process in interview, participants often described making separate notes on each of their readings, then transforming these into one summary or plan, which they then transform into an essay. This pipeline transforms representations A_1, A_2, \dots, A_n to B then B to C (Fig. 6ii). Other more complex pipelines may also be possible, such as a series of transformations where the final one draws from all previous representations: A to B , then B to C , then A, B and C to D (Fig. 6iii). Some pipelines might include translations that involve intensively updating the old representation while translating to the new one. These complex

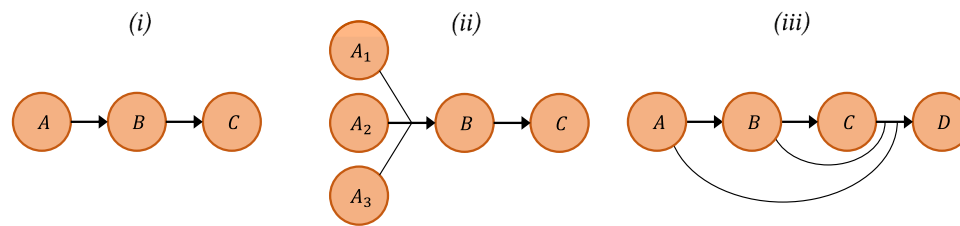


Fig. 6. Multiple transformation pipelines. We observed (i) in participant data. (ii) and (iii) are examples of possible, more complex pipelines.

pipelines might be beneficial for the essay-writing task, but only enabled by interfaces that support more sophisticated inter-representational interaction and some degree of automatic housekeeping of the relationship between the different representations.

These considerations are likely to apply beyond the domain of essay-writing. For example, in the domain of argument visualisation and mapping applications. Tools for argument mapping such as AGORA-net (Hoffmann, 2015), Compendium (Shum et al., 2006), Rationale van Gelder (2007), and others presented in the Literature Review (Section 2.2.1) often support the construction of a specialised type of argument map, and some are complemented with additional automatic tools such as tools to evaluate the conclusions (e.g., Cerutti et al., 2018; Janier et al., 2014). Similar to the essay-writing process, supporting pipelines of transformation between representations may be helpful in enabling access to these more specialised diagrams and their supporting tools.

Work in distributed cognition (e.g. Hutchins and Klausen, 1994; Zhang and Patel, 2006) analyses the propagation of representational state across different people's mental representations, speech, and physical structures. We highlight that, in the domain of essay-writing, these shifts between representational media are not only a propagation of state, but a generative transformation of it, in which ideas and meaning are developed. Future work can characterise the properties of these transformation pipelines between multiple people, and between permanent media and ephemeral media such as speech. Future work can also explore how transformations unfold in sensemaking activities, where analysts' evolving understanding of the data requires changes to the representational schema (Russell et al., 1993).

There is a rich possibility space here for future work to identify different pipelines, the kinds of representations, transformations and domains they are useful for, and tools to support them.

7.6. Conceptual distinctions in representational transformations

Zooming out, we now discuss the broader space of transformations. In our data, the 49 transformations we observed were all translations: they were not in-place because the second representation was created side-by-side with the first. In other contexts, and with other tools, *in-place* transformations may be more common. For example, an essay-writer using a Word processor might create a bullet-point list of topics, then gradually expand each bullet-point into a full section of an essay. Another example is a diagram that starts as a map of the relationships between ideas and, through pruning nodes and links, ends up as a graphical outline of the structure of an essay. The key difference is that after a translation, two different external representations exist, whereas after an in-place transformation, the new representation has replaced or overwritten the old. We believe there are important topics for future research relating to the permanence or transience of representations, and the trade-offs between translations and in-place transformations. For example, using translations means that the writer can refer back to previous representations, but this might become tedious or distracting if there are too many representations or when they start to diverge in content. Additionally, the properties of translations we identified (cardinality, explicitness, change in representation type) may also apply to in-place transformations, while their effects might differ.

In this paper, we have proposed a newly unified set of labels for

moving between representations (transformation, translation and transliteration), as we did not find an existing consistent use of terms in the literature, although we tried to maintain compatibility with the few existing uses of the terms. McCracken and Newstetter (2001) use 'transformation' to describe engineering students' move from problem description to diagram, noting that this is an important problem-solving skill. Keig and Rubba (1993) use 'translation' to describe chemistry students' methodical move between chemical formulas and ball-and-stick diagrams. We broaden these definitions while maintaining compatibility, defining a transformation as a representation *A* informing the creation of a representation *B*, and a translation as a transformation that is not in-place. This reflects that after a translation both representations still exist (just as translating a book to another language does not destroy or replace the original text). We think that these terms could be useful for UI designers and researchers in this area, as they have helped us conceptualise and discuss the different types of transformations and their properties. We look forward to the discussion of further subcategorisations that may be useful in other domains, and note that in other contexts it might be difficult to distinguish between some of our categories.

7.7. Lessons for practitioners

The starting point for our research was the design, implementation and testing of the tool, Write Reason. This tool was a valuable platform to investigate how essay writers use diverse representations. Nevertheless, reflecting on the design of the tool can be instructive. It is now clear to us that the tool embedded assumptions about the essay-writing process that our observations exposed as naïve or simplistic. Highlighting these might help designers of similar kinds of systems start one step ahead in their designs. More specifically, Write Reason:

- Assumed a limited number of representations (an essay and a map)
- Assumed that relationships between items in separate representations would be simple (e.g., mostly one-to-one in cardinality, with similar levels of explicitness)
- Overlooked the need to add meta annotations and to track the provenance of items

Taking these considerations, the results of our studies, and the existing literature into account, we extract the following lessons for practitioners seeking to implement multi-representational tools for essay-writing:

- Writers are likely to have varying styles of creating representations; unless a normative approach is the goal (e.g., following a specific essay-writing methodology), the representation interface should be highly flexible and enable multiple ways to relate items to each other (e.g., labelled relationships, provenance).
- Many writers want the ability to create multiple representations, often beyond just two (e.g., an intermediate conceptual representation and the final output text itself); consider designing interfaces that are not limited in the number and type of representations that can be created.

- The processes by which information flows between representations are anything but simple; consider that important cognitive activity takes place precisely in the translation between representations.

Some of these lessons might apply more broadly to multi-representational interfaces for other tasks, but further work is required for these to generalise to other domains. Additionally, how best to support the creation and maintenance of relationships between multiple representations is an open question. Further studies are required to establish the benefits and drawbacks of automatic or manual translations for different complex and cognitively demanding tasks.

7.8. Limitations

The work presented in this article represents an early step towards knowledge to support better essay-writing through technology and an improved understanding of the role of external representations and transformations between representations in complex tasks. As such, it is necessarily incomplete, imperfect and limited. Here we bring the reader's attention to its limitations to further qualify and support the interpretation of our contributions.

First, our qualitative observational approach implies that the evidence that we gathered is correlational and might not hold if conditions are explicitly manipulated. For example, although we observed a negative correlation between the use of one-to-one transformations and essay scores, we do not know if forcing writers to use one-to-one transformations would result in degraded essay scores. In particular, more studies are needed to understand whether the correlations we observed may be affected by differences in background essay-writing skill, patterns of thought, and mapping familiarity.

Second, the number of participants, although appropriate for a study of these characteristics, does not allow for sweeping generalisations, which we have avoided. Further studies that isolate specific effects are necessary to ascertain that the effects are stable. Closely related to this, our participant sample is most representative of university students in the Western world. Generalising the results to other cultures and ages will require additional work.

Third, the study and its observations are necessarily influenced by the specific design of the tool, Write Reason. It is possible that a different tool with a different design could have resulted in somewhat different observations. Nevertheless, we think that it is unlikely that the key findings from our study (e.g., the variety and complexity of translations) are entirely an artefact of Write Reason features. It is more likely that different designs could partially influence which types or characteristics of translations are more common, but not that the important cognitive processing observed taking place through translations would disappear. Replications of this experiment using a different design, or even a fully paper-based workflow could be useful to confirm and generalise the findings.

Fourth, we only exposed participants to representations in the form of graphical diagrams and text. Although these likely cover a large proportion of the representations that essay writers would consider using, there are other possibilities in other types of media and with different affordances that some people might use or that support essay-writing for specific populations. Consider, for example, audio memo notes, or sticky note diagrams.

Finally, throughout this work we have deliberately constrained ourselves to external representations. Hence the results are focused on markings and actions that can be recorded by the system, or are reported by participants. Internal representations are also highly important, but much more difficult to access because they are stored in the head of the writer. Insofar as the external representations provide evidence of cognitive processes, in the distributed cognition (Hollan et al., 2000) or embedded cognition (Clark, 2008) sense, we have tried to present and discuss this evidence. It is important to highlight, however, that further progress might depend on explicitly modelling how internal

representations interact with external representations in the process of essay-writing and other tasks, likely requiring measurements and methods beyond those that we applied.

7.9. Open questions for future research

Although software support for writing and the study of representations in complex cognitive tasks are long established research areas, we have observed a recent resurgence of this type of research (e.g., Barstow et al., 2017b; Fan et al., 2019; Harrell and Wetzel, 2015; Jafari and Zarei, 2015; Zhu et al., 2020). We believe that substantial benefits are to be reaped from increased knowledge in this area, in the form of software that supports the complex cognitive processes of students and knowledge workers in better ways. But this requires answers to questions that we have only started answering. The following are a small selection, starting with those questions for which we have not been able to find satisfactory answers (e.g., Q3 above):

- Which of the representation and translation practices that we observed improve essay quality?
- Do different types of transformations between representations support different types of cognitive processes (e.g., convergent vs. divergent thinking)?
- Can training writers to follow specific transformation processes improve their writing?
- Will support from software (e.g., automatic synchronisation of multiple representations) give writers access to workflows that are beneficial but complex?
- What additional properties of transformations are observed in other domains, timescales and populations?

8. Conclusion

External representations can be powerful aids for complex tasks like essay-writing (Larkin and Simon, 1987; van Gelder, 2007; Zhang and Norman, 1994). The affordances of intermediate representations, such as concept maps and argument maps, are relatively well understood (Davies, 2011; Novak and Cañas, 2006; Reed et al., 2007), allowing the development of tools to support their construction (e.g., Badam et al., 2019; Cañas et al., 2004; Carneiro et al., 2019; Introne, 2009; Wang et al., 2019). However, in this paper, we have argued that how people use these representations in the essay-writing workflow is an important yet neglected topic, and understanding it is crucial to develop effective multi-representational tools.

We presented the findings of a study, using mostly qualitative analysis that examines the representations and processes of 20 essay-writers. Our tool, Write Reason, offered insight into how people created, used, and moved between map and text representations. This led us to develop the concept of representational transformations. We found that participants universally used translations, rather than in-place transformations. Participants mostly used batch translations, building an entire representation (e.g., a map of the evidence) then translating it to another (e.g., an essay text). Participants often built more than just a single map and text representation, using Write Reason flexibly to perform translations with different cardinalities (one-to-one or one-to-many), and changes in explicitness.

We have highlighted implications of these findings, and open questions that remain. These sketch some rough contours of a large design space of tools to support representational transformations in complex tasks. By improving our understanding of how people use intermediate representations in the essay-writing workflow, we hope to support the design of a new generation of effective multi-representational tools for essay-writing and other complex tasks.

CRediT authorship contribution statement

Adam Binks: Conceptualization, Software, Investigation, Writing – original draft. **Alice Toniolo:** Conceptualization, Validation, Writing – review & editing. **Miguel A. Nacenta:** Conceptualization, Validation, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This research was supported by NSERC (The Natural Sciences and Engineering Research Council of Canada) RGPIN-2020-04401 and EPSRC (Engineering and Physical Sciences Research Council) EP/T518062/1. We thank our reviewers and other colleagues for comments that greatly improved the manuscript.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:[10.1016/j.ijhcs.2022.102851](https://doi.org/10.1016/j.ijhcs.2022.102851)

References

- Azlinda, W., Mohamed, W., Omar, B., 2008. Using concept map to facilitate writing assignment. *Proceedings of the Third International Conference on Concept Mapping*, Vol. 2, pp. 5–9.
- Badam, S.K., Liu, Z., Elmquist, N., 2019. Elastic documents: coupling text and tables through contextual visualizations for enhanced document reading. *IEEE Trans. Vis. Comput. Graph.* 25 (1), 661–671. <https://doi.org/10.1109/TVCG.2018.2865119>.
- Bar, H., Mentch, L., 2017. R-CMap—an open-source software for concept mapping. *Eval. Program Plann.* 60, 284–292. <https://doi.org/10.1016/j.evalprogplan.2016.08.018>.
- Barral, O., Lallé, S., Conati, C., 2020. Understanding the effectiveness of adaptive guidance for narrative visualization: a gaze-based analysis. *Proceedings of the 25th International Conference on Intelligent User Interfaces*. Association for Computing Machinery, New York, NY, USA, pp. 1–9. <https://doi.org/10.1145/3377325.3377517>.
- Barstow, B., Fazio, L., Lippman, J., Falakmasir, M., Schunn, C.D., Ashley, K.D., 2017. The impacts of domain-general vs. domain-specific diagramming tools on writing. *Int. J. Artif. Intell. Educ.* 27 (4), 671–693. <https://doi.org/10.1007/s40593-016-0130-z>.
- Barstow, B., Fazio, L., Schunn, C., Ashley, K., 2017. Experimental evidence for diagramming benefits in science writing. *Instr. Sci.* 45 (5), 537–556. <https://doi.org/10.1007/s11251-017-9415-3>.
- Barzilai, S., Mor-Hagani, S., Zohar, A.R., Shlomi-Elouz, T., Ben-Yishai, R., 2020. Making sources visible: promoting multiple document literacy with digital epistemic scaffolds. *Comput. Educ.* 157, 103980. <https://doi.org/10.1016/j.compedu.2020.103980>.
- Barzilai, S., Tal-Savir, D., Abed, F., Mor-Hagani, S., Zohar, A.R., 2021. Mapping multiple documents: from constructing multiple document models to argumentative writing. *Read. Writ.* <https://doi.org/10.1007/s11145-021-10208-8>.
- Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc.* 57 (1), 289–300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>.
- Biggs, J., 1988. Approaches to learning and to essay writing. In: Schmeck, R.R. (Ed.), *Learning Strategies and Learning Styles, Perspectives on Individual Differences*. Springer US, Boston, MA, pp. 185–228. https://doi.org/10.1007/978-1-4899-2118-5_8.
- Blackwell, A.F., 2002. Psychological perspectives on diagrams and their users. In: Anderson, M., Meyer, B., Olivier, P. (Eds.), *Diagrammatic Representation and Reasoning*. Springer, London, pp. 109–123. https://doi.org/10.1007/978-1-4471-0109-3_6.
- Buja, A., McDonald, J.A., Michalak, J., Stuetzle, W., 1991. Interactive data visualization using focusing and linking. *Proceedings of the Second IEEE Conference on Visualization*. IEEE, New York, NY, USA, pp. 156–163. <https://doi.org/10.1109/VISUAL.1991.175794>.
- Butchart, S., Forster, D., Gold, I., Bigelow, J., Korb, K., Oppy, G., Serrenti, A., 2009. Improving critical thinking using web based argument mapping exercises with automated feedback. *Aust. J. Educ. Technol.* 25 (2), 268–291. <https://doi.org/10.14742/ajet.1154>.
- Butcher, K.R., 2014. The multimedia principle. *The Cambridge Handbook of Multimedia Learning*, 2nd ed. In: Cambridge Handbooks in Psychology. Cambridge University Press, New York, NY, US, pp. 174–205. <https://doi.org/10.1017/CBO9781139547369.010>.
- Campbell, J., Smith, D., Brooker, R., 1998. From conception to performance: how undergraduate students conceptualise and construct essays. *High. Educ.* 36 (4), 449–469. <https://doi.org/10.1023/A:1003451627898>.
- Canas, A.J., Hill, G., Carff, R., Suri, N., Lott, J., Gómez, G., Eskridge, T.C., Arroyo, M., Carvajal, R., 2004. CmapTools: a knowledge modeling and sharing environment. *Proceedings of the First International Conference on Concept Mapping*, Vol. 1, pp. 125–133. Pamplona, Spain.
- Carneiro, G., Nacenta, M., Toniolo, A., Mendez, G., Quigley, A.J., 2019. Deb8: a tool for collaborative analysis of video. *Proceedings of the 2019 ACM International Conference on Interactive Experiences for TV and Online Video*. Association for Computing Machinery, New York, NY, USA, pp. 47–58. <https://doi.org/10.1145/3317697.3323358>.
- Carroll, J.M., Wu, Y., Shih, P.C., Zheng, S., 2016. Re-appropriating a question/answer system to support dialectical constructivist learning activity. *Educ. Technol. Res. Dev.* 64 (1), 137–156. <https://doi.org/10.1007/s11423-015-9405-6>.
- Castro-Alonso, J.C., Ayres, P., Sweller, J., 2019. Instructional visualizations, cognitive load theory, and visuospatial processing. In: Castro-Alonso, J.C. (Ed.), *Visuospatial Processing for Education in Health and Natural Sciences*. Springer, New York, NY, USA, pp. 111–143. https://doi.org/10.1007/978-3-030-20969-8_5.
- Cerutti, F., Gaggl, S.A., Thimm, M., Wallner, J.P., 2017. Foundations of implementations for formal argumentation. *IfCoLog J. LogicsAppl. Spec. Issue Formal Argumentation* 4, 2623–2705.
- Cerutti, F., Norman, T. J., Toniolo, A., Middleton, S., 2018. CISpaces.org: from fact extraction to report generation information.
- Chan, C.C., Tsui, M.S., Chan, M.Y.C., Hong, J.H., 2002. Applying the structure of the observed learning outcomes (SOLO) taxonomy on student's learning outcomes: an empirical study. *Asses. Eval. Higher Educ.* 27 (6), 511–527. <https://doi.org/10.1080/0260293022000020282>.
- Cherubini, M., Venolia, G., DeLine, R., Ko, A.J., 2007. Let's go to the whiteboard: how and why software developers use drawings. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, pp. 557–566. <https://doi.org/10.1145/1240624.1240714>.
- Chiang, K.-H., Fan, C.-Y., Liu, H.-H., Chen, G.-D., 2016. Effects of a computer-assisted argument map learning strategy on sixth-grade students' argumentative essay reading comprehension. *Multimed Tools Appl.* 75 (16), 9973–9990. <https://doi.org/10.1007/s11042-015-2904-y>.
- Clark, A., 2008. *Supersizing the Mind: Embodiment, Action, and Cognitive Extension*. In: *Philosophy of Mind Series*. Oxford University Press, New York. <https://doi.org/10.1093/acprof:oso/9780195333213.001.0001>.
- Corbin, J., Strauss, A., 2008. *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory*, 3rd Ed. In: Sage Publications, Thousand Oaks, CA, US.
- Cropley, A., 2006. In praise of convergent thinking. *Creat. Res. J.* 18 (3), 391–404. https://doi.org/10.1207/s15326934crj1803_13.
- Cullen, S., Fan, J., van der Brugge, E., Elga, A., 2018. Improving analytical reasoning and argument understanding: a quasi-experimental field study of argument visualization. *npj Sci. Learn.* 3 (1), 1–6. <https://doi.org/10.1038/s41539-018-0038-5>.
- Davies, M., 2011. Concept mapping, mind mapping and argument mapping: what are the differences and do they matter? *High. Educ.* 62 (3), 279–301. <https://doi.org/10.1007/s10734-010-9387-6>.
- de Vries, E., 2012. Learning with external representations. In: Seel, N.M. (Ed.), *Encyclopedia of the Sciences of Learning*. Springer US, Boston, MA, pp. 2016–2019. https://doi.org/10.1007/978-1-4419-1428-6_675.
- Dufresne, R.J., Gerace, W.J., Leonard, W.J., 1997. Solving physics problems with multiple representations. *Phys. Teacher* 35 (5), 270–275. <https://doi.org/10.1119/1.2344681>.
- 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 Creativity* 5 (1), 16–22. <https://doi.org/10.1016/j.tsc.2009.05.001>.
- Easterday, M.W., Alevan, V., Scheines, R., Carver, S.M., 2009. Constructing causal diagrams to learn deliberation. *Int. J. Artif. Intell. Educ.* 19 (4), 425–445.
- Eklundh, K.S., 1992. Problems in achieving a global perspective of the text in computer-based writing. In: Sharples, M. (Ed.), *Computers and Writing: Issues and Implementations*. Springer Netherlands, Dordrecht, pp. 73–84. https://doi.org/10.1007/978-94-011-2674-8_6.
- Eppler, M.J., 2006. A comparison between concept maps, mind maps, conceptual diagrams, and visual metaphors as complementary tools for knowledge construction and sharing. *Inf. Vis.* 5 (3), 202–210. <https://doi.org/10.1057/palgrave.ivs.9500131>.
- Fan, C.-Y., Liyanawatta, M., Yang, S.-H., Chen, G.-D., 2019. Using digital map tools to assist learning of argumentative essay writing. In: Rønningsbakk, L., Wu, T.-T., Sandnes, F.E., Huang, Y.-M. (Eds.), *Innovative technologies and learning*. Springer, New York, NY, USA, pp. 794–801. https://doi.org/10.1007/978-3-030-35343-8_82.
- Furniss, D., Blandford, A., Curzon, P., 2011. Confessions from a grounded theory PhD: experiences and lessons learnt. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, pp. 113–122. <https://doi.org/10.1145/1978942.1978960>.
- Grossen, B., Carmine, D., 1990. Diagramming a logic strategy: effects on difficult problem types and transfer. *Learn. Disabil. Q.* 13 (3), 168–182. <https://doi.org/10.2307/1510699>.
- Harrell, M., Wetzel, D., 2013. Improving first-year writing using argument diagramming. *Proceedings of the Annual Meeting of the Cognitive Science Society*, vol. 35, pp. 2488–2493.
- Harrell, M., Wetzel, D., 2015. Using argument diagramming to teach critical thinking in a first-year writing course. In: Davies, M., Barnett, R. (Eds.), *The Palgrave Handbook*

- of Critical Thinking in Higher Education. Palgrave Macmillan US, New York, pp. 213–232. https://doi.org/10.1057/9781137378057_14.
- Hayes, J.R., Flower, L.S., 1987. Writing research and the writer. *Am. Psychol.* 41 (10), 1106. <https://doi.org/10.1037/0003-066X.41.10.1106>.
- Hegarty, M., 2004. Diagrams in the mind and in the world: relations between internal and external visualizations. In: Blackwell, A.F., Marriott, K., Shimojima, A. (Eds.), *Diagrammatic Representation and Inference, Lecture Notes in Computer Science*. Springer Berlin Heidelberg, pp. 1–13.
- Hoffmann, M.H.G., 2015. Changing philosophy through technology: complexity and computer-supported collaborative argument mapping. *Philos. Technol.* 28 (2), 167–188. <https://doi.org/10.1007/s13347-013-0143-6>.
- Hollan, J., Hutchins, E., Kirsh, D., 2000. Distributed cognition: toward a new foundation for human-computer interaction research. *ACM Trans. Comput.-Hum. Interact.* 7 (2), 174–196. <https://doi.org/10.1145/353485.353487>.
- Hutchins, E., Klausen, T., 1994. Distributed cognition in an airline cockpit. *Communication and Cognition at Work*.
- Introne, J.E., 2009. Supporting group decisions by mediating deliberation to improve information pooling. Proceedings of the ACM 2009 International Conference on Supporting Group Work. Association for Computing Machinery, New York, NY, USA, pp. 189–198. <https://doi.org/10.1145/1531674.1531702>.
- Jafari, N., Zarei, G.R., 2015. The influence of concept mapping on iranian intermediate EFL learners' argumentative essay writing skill. *J. Appl. Linguist. Lang. Res.* 2 (4), 98–112.
- Janier, M., Lawrence, J., Reed, C., 2014. OVA+: an argument analysis interface. *Computational Models of Argument: Proceedings of COMMA 2014*. IOS Press, Amsterdam, pp. 463–464. <https://doi.org/10.3233/978-1-61499-436-7-463>.
- Keig, P.F., Rubba, P.A., 1993. Translation of representations of the structure of matter and its relationship to reasoning, gender, spatial reasoning, and specific prior knowledge. *J. Res. Sci. Teach.* 30 (8), 883–903. <https://doi.org/10.1002/tea.3660300807>.
- Kinneavy, J.L., 1971. *A Theory of Discourse: The Aims of Discourse*. Prentice-Hall, Englewood Cliffs, NJ.
- Kirsh, D., Maglio, P., 1994. On distinguishing epistemic from pragmatic action. *Cogn. Sci.* 18 (4), 513–549. [https://doi.org/10.1016/0364-0213\(94\)90007-8](https://doi.org/10.1016/0364-0213(94)90007-8).
- Klein, M., 2012. Enabling large-scale deliberation using attention-mediation metrics. *Comput. Support. Coop. Work (CSCW)* 21 (4), 449–473. <https://doi.org/10.1007/s10606-012-9156-4>.
- Kleinberg, J., Tardos, E., 2006. *Algorithm Design*. Pearson Education India.
- Kohnle, A., Passante, G., 2017. Characterizing representational learning: a combined simulation and tutorial on perturbation theory. *Phys. Rev. Phys. Educ. Res.* 13 (2), 020131. <https://doi.org/10.1103/PhysRevPhysEducRes.13.020131>.
- Kong, N., Hearst, M.A., Agrawala, M., 2014. Extracting references between text and charts via crowdsourcing. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. Association for Computing Machinery, New York, NY, USA, pp. 31–40. <https://doi.org/10.1145/2556288.2557241>.
- Korf, R.E., 1985. Depth-first iterative-deepening: an optimal admissible tree search. *Artif. Intell.* 27 (1), 97–109.
- Kozma, R., Russell, J., 2005. Students becoming chemists: developing representational competence. In: Gilbert, J.K. (Ed.), *Visualization in Science Education, Models and Modeling in Science Education*. Springer Netherlands, Dordrecht, pp. 121–145. https://doi.org/10.1007/1-4020-3613-2_8.
- Kumar, A., Saigal, R., 2005. Visual understanding environment. Proceedings of the 5th ACM/IEEE-CS Joint Conference on Digital Libraries. Association for Computing Machinery, New York, NY, USA, p. 413. <https://doi.org/10.1145/1065385.1065517>.
- Kunz, W., Rittel, H.W.J., 1970. *Issues as Elements of Information Systems*. Institute of Urban and Regional Development, University of California, Berkeley.
- Kwon, B. C., Stoffel, F., Jäckle, D., Lee, B., Keim, D., 2014. VisJockey: enriching data stories through orchestrated interactive visualization.
- Lai, C., Lin, Z., Jiang, R., Han, Y., Liu, C., Yuan, X., 2020. Automatic annotation synchronizing with textual description for visualization. Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. Association for Computing Machinery, New York, NY, USA, pp. 1–13. <https://doi.org/10.1145/3313831.3376443>.
- Lallé, S., Toker, D., Conati, C., 2019. Gaze-driven adaptive interventions for magazine-like narrative visualizations. *IEEE Transactions on Visualization and Computer Graphics*, pp. 1–12. <https://doi.org/10.1109/TVCG.2019.2958540>. Early Access
- Larkin, J.H., Simon, H.A., 1987. Why a diagram is (sometimes) worth ten thousand words. *Cogn. Sci.* 11 (1), 65–100. [https://doi.org/10.1016/S0364-0213\(87\)80026-5](https://doi.org/10.1016/S0364-0213(87)80026-5).
- Lavelle, E., Zuercher, N., 2001. The writing approaches of university students. *High. Educ.* 42 (3), 373–391. <https://doi.org/10.1023/A:1017967314724>.
- Liu, Y.-F., Fan, C.-Y., Chang, Y.-H., Chen, G.-D., 2017. A digital tool for argumentation construction that assists users in writing argumentative essays. 2017 IEEE 17th International Conference on Advanced Learning Technologies (ICALT), pp. 411–413. <https://doi.org/10.1109/ICALT.2017.15>.
- Lohse, G.L., Biolsi, K., Walker, N., Rueter, H.H., 1994. A classification of visual representations. *Commun. ACM* 37 (12), 36–49. <https://doi.org/10.1145/198366.198376>.
- Lynch, C.F., Ashley, K.D., Chi, M., 2014. Can diagrams predict essay grades? In: Trausan-Matu, S., Boyer, K.E., Crosby, M., Panourgia, K. (Eds.), *Intelligent Tutoring Systems*. Springer, New York, NY, USA, pp. 260–265. https://doi.org/10.1007/978-3-319-07221-0_32.
- Mayer, R.E., 2002. *Multimedia learning. Psychology of Learning and Motivation, Vol. 41*. Academic Press, Amsterdam, pp. 85–139.
- McCracken, W.M., Newstetter, W.C., 2001. Text to diagram to symbol: representational transformations in problem-solving. Proceedings of the 31st Annual Frontiers in Education Conference. IEEE, New York, NY, USA, pp. F2G–13. <https://doi.org/10.1109/FIE.2001.963721>.
- McGinley, W., 1992. The role of reading and writing while composing from sources. *Read. Res. Q.* 27 (3), 227–248. <https://doi.org/10.2307/747793>.
- Moreno, R., Mayer, R.E., 1999. Cognitive principles of multimedia learning: the role of modality and contiguity. *J. Educ. Psychol.* 91 (2), 358–368. <https://doi.org/10.1037/0022-0663.91.2.358>.
- Nardi, B.A., 1993. *A Small Matter of Programming: Perspectives on End User Computing*. MIT press.
- Nesbit, J., Niu, H., Liu, Q., 2019. Cognitive tools for scaffolding argumentation. In: Adesope, O.O., Rud, A. (Eds.), *Contemporary Technologies in Education: Maximizing Student Engagement, Motivation, and Learning*. Springer International Publishing, Cham, pp. 97–117. https://doi.org/10.1007/978-3-319-89680-9_6.
- Neuwirth, C.M., Kaufer, D.S., 1989. The role of external representation in the writing process: implications for the design of hypertext-based writing tools. Proceedings of the Second Annual ACM Conference on Hypertext. Association for Computing Machinery, New York, NY, USA, pp. 319–341. <https://doi.org/10.1145/74224.74250>.
- Newell, G.E., Beach, R., Smith, J., VanDerHeide, J., 2011. Teaching and learning argumentative reading and writing: a review of research. *Read. Res. Q.* 46 (3), 273–304. <https://doi.org/10.1598/RRQ.46.3.4>.
- Novak, J., Cañas, A., 2006. *The Theory Underlying Concept Maps and How to Construct Them*. Technical Report. Florida Institute for Human and Machine Cognition, Pensacola, FL.
- Nussbaum, E.M., Schraw, G., 2007. Promoting argument-counterargument integration in students' writing. *J. Exp. Educ.* 76 (1), 59–92. <https://doi.org/10.3200/JEXE.76.1.59-92>.
- Okada, A., 2008. Scaffolding School Pupils' Scientific Argumentation with Evidence-based Dialogue Maps. In: Okada, A., Shum, S.B., Sherborne, T. (Eds.), *Knowledge cartography: software tools and mapping techniques, Advanced Information and Knowledge Processing*. Springer, London, pp. 131–162. https://doi.org/10.1007/978-1-84800-149-7_7.
- Paas, F., Sweller, J., 2014. Implications of cognitive load theory for multimedia learning. *The Cambridge Handbook of Multimedia Learning, 2nd ed.* In: Cambridge Handbooks in Psychology. Cambridge University Press, New York, NY, US, pp. 27–42. <https://doi.org/10.1017/CBO9781139547369.004>.
- Plotnick, E., 1997. *Concept Mapping: A Graphical System for Understanding the Relationship between Concepts*. ERIC Digest. ERIC, Syracuse, NY.
- Prosser, M., Webb, C., 1994. Relating the process of undergraduate essay writing to the finished product. *Stud. High. Educ.* 19 (2), 125–138. <https://doi.org/10.1080/03075079412331381987>.
- Reddy, Y.M., Andrade, H., 2010. A review of rubric use in higher education. *Asses. Eval. Higher Educ.* 35 (4), 435–448. <https://doi.org/10.1080/02602930902862859>.
- Reed, C., Walton, D., Macagno, F., 2007. Argument diagramming in logic, law and artificial intelligence. *Knowl. Eng. Rev.* 22 (1), 87–109.
- Rolf, B., Magnusson, C., 2002. Developing the art of argumentation - a software approach. Proceedings of the 5th International Conference on Argumentation. International Society for the Study of Argumentation, Amsterdam.
- Runco, M.A., Acar, S., 2012. Divergent thinking as an indicator of creative potential. *Creat. Res. J.* 24 (1), 66–75. <https://doi.org/10.1080/10400419.2012.652929>.
- Russell, D.M., Stefik, M.J., Pirolli, P., Card, S.K., 1993. The cost structure of sensemaking. Proceedings of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems. Association for Computing Machinery, New York, NY, USA, pp. 269–276. <https://doi.org/10.1145/169059.169209>.
- Scaife, M., Rogers, Y., 1996. External cognition: how do graphical representations work? *Int. J. Hum. Comput. Stud.* 45 (2), 185–213. <https://doi.org/10.1006/ijhc.1996.0048>.
- Schön, D.A., 1992. Designing as reflective conversation with the materials of a design situation. *Knowl. Based Syst.* 5 (1), 3–14. [https://doi.org/10.1016/0950-7051\(92\)90020-G](https://doi.org/10.1016/0950-7051(92)90020-G).
- Schuler, V., Smith, J.B., 1992. Author's argumentation assistant (AAA): a hypertext-based authoring tool for argumentative texts. *Hypertext: Concepts, Systems and Applications*. Cambridge University Press, USA, pp. 137–151.
- Schwarz, B.B., Glassner, A., 2007. The role of floor control and of ontology in argumentative activities with discussion-based tools. *Int. J. Comput.-Supported Collab. Learn.* 2 (4), 449–478. <https://doi.org/10.1007/s11412-007-9024-2>.
- Selvin, A., Buckingham Shum, S., Seirhuis, M., Conklin, J., Zimmerman, B., Palus, C., Drath, W., Horth, D., Domingue, J., Motta, E., Li, G., 2001. *Compendium: Making meetings into knowledge events*. Knowledge Technologies 2001. Texas, USA
- Shum, S.J.B., Selvin, A.M., Sierhuis, M., Conklin, J., Haley, C.B., Nuseibeh, B., 2006. Hypermedia support for argumentation-based rationale. In: Dutoit, A.H., McCall, R., Mistrik, I., Paech, B. (Eds.), *Rationale Management in Software Engineering*. Springer, Berlin, Heidelberg, pp. 111–132. https://doi.org/10.1007/978-3-540-30998-7_5.
- Simon, H.A., 1978. On the forms of mental representation. In: Savage, C.W. (Ed.), *Perception and Cognition: Issues in the Foundations of Psychology*. University of Minnesota Press, Minneapolis, MN, pp. 3–17.
- Slotte, V., Lonka, K., 2001. Note taking and essay writing. In: Tynjälä, P., Mason, L., Lonka, K. (Eds.), *Writing as a Learning Tool: Integrating Theory and Practice, Studies in Writing*. Springer Netherlands, Dordrecht, pp. 131–143.
- Smith, D., Campbell, J., Brooker, R., 1999. The impact of students' approaches to essay writing on the quality of their essays. *Asses. Eval. Higher Educ.* 24 (3), 327–338. <https://doi.org/10.1080/0260293990240306>.

- Smith, J.B., Weiss, S.F., Ferguson, G.J., 1987. A hypertext writing environment and its cognitive basis (panel session). Proceedings of the ACM Conference on Hypertext. Association for Computing Machinery, New York, NY, USA, pp. 195–214. <https://doi.org/10.1145/317426.317442>.
- Sorden, S.D., 2012. The cognitive theory of multimedia learning. *Handb. Educ. Theor.*
- Spivey, N.N., 1990. Transforming texts: constructive processes in reading and writing. *Written Commun.* 7 (2), 256–287. <https://doi.org/10.1177/0741088390007002004>.
- Streitz, N.A., Hannemann, J., Thüring, M., 1989. From ideas and arguments to hyperdocuments: travelling through activity spaces. Proceedings of the Second Annual ACM Conference on Hypertext. Association for Computing Machinery, New York, NY, USA, pp. 343–364. <https://doi.org/10.1145/74224.74251>.
- Suthers, D., Weiner, A., Connelly, J., Paolucci, M., 1995. Belvedere: engaging students in critical discussion of science and public policy issues. Proceedings of the 7th World Conference on Artificial Intelligence in Education. International Artificial Intelligence in Education Society, Washington, DC, pp. 266–273.
- Sweller, J., 2011. Cognitive load theory. *The Psychology of Learning and Motivation: Cognition in Education*. In: *The Psychology of Learning and Motivation*, Vol. 55. Elsevier Academic Press, San Diego, CA, US, pp. 37–76. <https://doi.org/10.1016/B978-0-12-387691-1.00002-8>.
- Toniolo, A., Norman, T.J., Etuk, A., Cerutti, F., Ouyang, R.W., Srivastava, M., Oren, N., Dropps, T., Allen, J.A., Sullivan, P., 2015. Supporting reasoning with different types of evidence in intelligence analysis. Proceedings of the Fourteenth International Conference on Autonomous Agents and Multiagent Systems, pp. 781–789.
- Toulmin, S.E., 2003. *The Uses of Argument*. Cambridge University Press.
- Twardy, C., 2004. Argument maps improve critical thinking. *Teach. Philos.* 27 (2), 95–116. <https://doi.org/10.5840/teachphil200427213>.
- van den Braak, S.W., van Oostendorp, H., Prakken, H., Vreeswijk, G.a.W., 2008. A critical review of argument visualization tools: do users become better reasoners? Workshop Notes of the ECAI-06 Workshop on Computational Models of Natural Argument, pp. 67–75.
- van Gelder, T., 2002. Argument mapping with reason!able. *Am. Philos. Assoc. Newsl. Philos. Comput.* 2, 85–90.
- van Gelder, T., 2007. The rationale for RationaleTM. *Law Probab. Risk* 6 (1–4), 23–42. <https://doi.org/10.1093/lpr/mgm032>.
- Villalon, J., Calvo, R.A., 2011. Concept maps as cognitive visualizations of writing assignments. *J. Educ. Technol. Soc.* 14 (3), 16–27.
- Walny, J., Carpendale, S., Riche, N.H., Venolia, G., Fawcett, P., 2011. Visual thinking in action: visualizations as used on whiteboards. *IEEE Trans. Vis. Comput. Graph.* 17 (12), 2508–2517. <https://doi.org/10.1109/TVCG.2011.251>.
- Walton, D., 2005. *Fundamentals of Critical Argumentation*. Cambridge University Press.
- Wang, S., Sonmez Unal, D., Walker, E., 2019. MindDot: supporting effective cognitive behaviors in concept map-based learning environments. Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. Association for Computing Machinery, New York, NY, USA, pp. 1–14. <https://doi.org/10.1145/3290605.3300258>.
- Wong, C., 2007. Views on the adoption and implementation of the SOLO taxonomy. In: Frankland, S. (Ed.), *Enhancing Teaching and Learning through Assessment: Deriving an Appropriate Model*. Springer, New York, NY, pp. 4–15.
- Yamamoto, Y., Takada, S., Gross, M., Nakakoji, K., 1998. Representational talkback: an approach to support writing as design. Proceedings. 3rd Asia Pacific Computer Human Interaction (Cat. No.98EX110), pp. 125–131. <https://doi.org/10.1109/APCHI.1998.704176>.
- Zarina, P., Fatima, H., 2015. The effects of concept mapping on matriculation students' essay writing performance. Proceedings of the International Conference on Education and Social Sciences. International Organization Center of Academic Research, Istanbul, pp. 1326–1333.
- Zhang, J., 1997. The nature of external representations in problem solving. *Cogn. Sci.* 21 (2), 179–217. https://doi.org/10.1207/s15516709cog2102_3.
- Zhang, J., Norman, D.A., 1994. Representations in distributed cognitive tasks. *Cogn. Sci.* 18 (1), 87–122. https://doi.org/10.1207/s15516709cog1801_3.
- Zhang, J., Patel, V.L., 2006. Distributed cognition, representation, and affordance. *Pragmat. Cognit.* 14 (2), 333–341. <https://doi.org/10.1075/pc.14.2.12zha>.
- Zhang, P., Soergel, D., 2014. Towards a comprehensive model of the cognitive process and mechanisms of individual sensemaking. *J. Assoc. Inf. Sci. Technol.* 65 (9), 1733–1756. <https://doi.org/10.1002/asi.23125>.
- Zhi, Q., Ottley, A., Metoyer, R., 2019. Linking and layout: exploring the integration of text and visualization in storytelling. *Comput. Graphics Forum* 38 (3), 675–685. <https://doi.org/10.1111/cgf.13719>.
- Zhu, X., Nacenta, M.A., Akgün, Ö., Nightingale, P., 2020. How people visually represent discrete constraint problems. *IEEE Trans. Vis. Comput. Graph.* 26 (8), 2603–2619. <https://doi.org/10.1109/TVCG.2019.2895085>.